Charles University

Faculty of Science

Study programme: Demography

Branch of study: Demography

Bc. Vojtěch Horník

Health expectancy in Europe: analysis of explanatory factors

Zdravá délka života v Evropě: analýza vysvětlujících faktorů

Diploma thesis

Supervisor: RNDr. Klára Hulíková Tesárková, Ph.D.

Prague, 2020

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 17. 4. 2020 …………………………………

Podpis

Na tomto místě bych rád poděkoval vedoucí této diplomové práce, RNDr. Kláře Hulíkové Tesárkové, Ph.D., za všechny rady, připomínky a komentáře k předkládané práci, ochotu kdykoli pomoci a za čas strávený nad čtením průběžných verzí práce.

Zároveň bych rád vyjádřil vděk svým nejbližším za to, že pro mě představovali během tvorby této práce důležitou oporu. Ta byla potřeba zejména v době dokončování textu, kdy bylo v platnosti omezení volného pohybu, a prostor pro jakoukoli práci se tak omezil pouze na několik desítek čtverečních metrů mého domova. Nemohlo to být se mnou v tomto čase jednoduché.

Děkuji.

Health expectancy in Europe: analysis of explanatory factors

Abstract

Health expectancies are aggregate measures of morbidity and mortality. This thesis analyzes health expectancies in European countries, especially in relation to factors that explain variation in values of health expectancy. Healthy life years and healthy life expectancy based on self-perceived health are chosen as specific indicators used throughout the analysis. First, trends in these measures are analyzed in the period 2005–2017. Second, healthy life expectancy based on self-perceived health is regressed on gross domestic product (GDP), measures of education, long-term unemployment rate, share of government expenditure on health and a measure of income inequality. Income inequality and education are found to have significant associations with health expectancy. For GDP and health expenditures, results were not conclusive. Unemployment was found to be associated with health expectancy positively, contradictory to expectations.

Keywords: health expectancy, healthy life years, self-perceived health, regression, morbidity

Zdravá délka života v Evropě: analýza vysvětlujících faktorů

Abstrakt

Ukazatele zdravé délky života v sobě kombinují mortalitu a morbiditu. Tato práce analyzuje zdravou délku života v evropských státech, a to zejména v souvislosti s faktory, které vysvětlují její variaci. Jako konkrétní indikátory použité v průběhu analýzy byly vybrány délka života ve zdraví (indikátor označovaný Eurostatem jako healthy life years) a naděje dožití ve zdraví podle subjektivního vnímání zdravotního stavu. Nejprve jsou analyzovány trendy v hodnotách těchto ukazatelů v období 2005–2017. Následně je provedena regrese naděje dožití ve zdraví podle subjektivního vnímání zdravotního stavu na hrubém domácím produktu (HDP), indikátorech vzdělání, míře dlouhodobé nezaměstnanosti, podílu vládních výdajů na zdravotnictví a na indikátoru příjmové nerovnosti. Příjmová nerovnost a vzdělání mají podle výsledků regrese signifikantní vztah s nadějí dožití ve zdraví. Výsledky pro HDP a výdaje na zdravotnictví nejsou jednoznačné. U nezaměstnanosti je odhadován pozitivní vztah s nadějí dožití ve zdraví, což neodpovídá očekáváním.

Klíčová slova: zdravá délka života, healthy life years, subjektivní vnímání zdraví, regrese, nemocnost

Number of characters without spaces / Počet znaků bez mezer: 111 522

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AROPE	at-risk-of-poverty or social exclusion
BLUE	best linear unbiased estimator
DALE	disability-adjusted life expectancy
DFLE	disability-free life expectancy
EU	European Union
EU-SILC	EU Statistics on Income and Living Conditions
FE	fixed effects
GALI	global activity limitation indicator
GBD	Global Burden of Disease
GDP	gross domestic product
HALE	health-adjusted life expectancy
HLE	healthy life expectancy
HLY	healthy life years
ISCED	international standardized classification of education
INED	Institut National d'Études Démographiques
LE	life expectancy
LFS	Labour Force Survey
OLS	ordinary least squares
RE	random effects
REVES	Réseau d'Espérance de Vie en Santé
SE	standard error
WHO	World Health Organization

Chapter 1

Introduction

People are living longer and longer. It is an undisputed fact that life expectancy has been increasing over the past decades in almost all of the world's countries; exceptions to this general trend are scarce. At the same time, most of the developed countries of the world that are undergoing or have already undergone the process of second demographic transition (as first described by Lesthaeghe and van de Kaa, 1986) experience low fertility. Their age structure therefore shifts and proportion of population in older ages increases.

With new challenges that ageing populations have to face, it became clear that apart from length of life itself, health of a population – mirrored afterwards in overall quality of life – is an aspect that should be of concern as well. Researchers started to be interested in a measure combining mortality (fatal outcomes) and morbidity (non-fatal outcomes) in 1960s (Sanders, 1964; Sullivan, 1966, 1971). This meant combining an information about population health and life expectancy, such measures are since then referred to as health expectancies. Because health expectancies are by design relatively easily comprehensible, they gradually started to make their way into demographic research and practice. Nowadays, they are widely acknowledged and in center of focus of institutions such as World Health Organization (WHO) or Eurostat.

A number of societies monitor the length of healthy life, and increase in years spent in good health is included in many goals and strategies formulated in a national or cross-national setting. In the United States, the *Healthy People* programme by National Center for Health Statistics is aimed at increasing „quality and years of healthy life" as well as „eliminating health disparities" (Park et al., 2008, p. 329). Eurostat included healthy life years in a set of structural indicators, which were designed to monitor the progress that European countries are making in order to achieve the goals formulated in *Lisbon strategy* and its successor, the *EU 2020 strategy* (Rychtaříková, 2006; Walburn, 2010). WHO monitors health expectancy in the Global Burden of Disease project; its purpose is to assess and compare level of health on a worldwide scale, and formulate strategies based on the results (Salomon et al., 2012).

One specific area where health expectancies could be implemented directly in policy making are pension reforms. Currently, in many developed countries, there are ongoing debates about raising pension age or, more generally, about ensuring that pension systems keep functioning as population is ageing. There are countries in Europe that have already tied future increases of the age of retirement with advances in life expectancy – for example Netherlands, Denmark, Italy or Spain (Holub et al., 2015). However, such efforts face criticism from some researchers, because advances in length of life do not necessarily mean that years lived in good health increase as well (e.g. Majer et al., 2011; Bronnum-Hansen et al., 2017). Fouweather et al. (2015) found out that in 2010, in about a third of countries in European Union (EU), a 50-year old person could not, on average, expect to live until official retirement age without activity limitation. Employing health expectancy in place of life expectancy is therefore a possible solution for policies towards pension system, since in such a case, level of population health would be also accounted for.

1.1 Objectives and further structure

It follows from previous paragraphs that health expectancies are an important aspect of current demographic debates. They are routinely computed and used, as was already stressed, but it nevertheless remains a topic that deserves further research.

This thesis focuses on health expectancies in the context of Europe, and their relationship to other characteristics from the fields of macro-economy, labour market, sociodemography and education. The indicators of health expectancy started to be computed by Eurostat for the majority of EU countries in 2005, and the latest available year, for which its values are currently available, is 2017. This means that already a solid time series of such data is available. Despite that, studies taking advantage of broad availability of these data are very scarce – if anything, associations between health expectancy and other factors are only analyzed in a cross-sectional framework (Jagger et al., 2008; Fouweather et al., 2015).

The ambition of this text is therefore to study the measures of health expectancy and uncover the extent to which health expectancy is influenced by macro-economic, educational and sociodemographic factors in a longitudinal framework. These factors are employed since they can be presumably controlled by governments in a more direct way than health status itself, and results of such an analysis can thus assist in policy formulation. The outlook is that since time series of more than ten years of data is available, the conclusions will be more informative compared to those that were drawn on the basis of a simple cross-sectional analysis.

All of this leads to formulation of the objectives for this thesis. The **first objective** is to explore and describe trends in measures of health expectancy in European countries from 2005 onwards, and the relation of health expectancy to life expectancy. The **second, crucial objective** is to explain variation in measures of health expectancy in European countries with the help of indicators from the fields of macro-economy, sociodemography, education and labour market; and to quantify relationships between them. As a consequence of these goals, specific hypotheses are constructed on the basis of literature research and are formulated at the end of chapter 3. The assessment of validity of the hypotheses is supposed to contribute to fulfillment of the objectives.

The first goal will be reached mostly by performing exploratory statistical analysis on the analyzed indicators. In order to meet the second objective, there are two tools that are used. Firstly, regression analysis is run with the data in order to quantify the associations between health expectancy and factors that are supposed to explain its variation. Secondly, potential pathways mediating the effects that one variable has on another one are discussed based on existing literature. Conclusions from both these steps are then combined in order to fulfill the second objective.

Structure of this thesis follows the formulation of the objectives, so that they can be met. The rest of the thesis is thus structured in the following way: Second chapter provides an overview of theoretical concepts and current practice that are relevant for studying health expectancies. Chapter 3 discusses some literature that has been written on the topic of links between population health and various factors that can explain inter-country differentials in levels of health measures. As was already mentioned, hypotheses about health expectancy and its relation to other factors are stated at the end of this chapter. Subsequently, fourth chapter presents data that are used in analytical part of the thesis. Fifth chapter is aimed at exploratory analysis of data on health expectancy, life expectancy and other factors that are supposed to potentially influence health expectancy. In chapter 6, attention is turned to technical issues concerning methods used for regression. All methodology is explained, described and discussed as for relevance of the specific techniques used. Chapter 7 reports results of the analysis and discusses their statistical and practical significance. Finally, chapter 8 concludes the thesis, sums up the findings, evaluates fulfillment of the objectives and discusses validity of the hypotheses.

Chapter 2

Theoretical background

This chapter aims to introduce health expectancies from a theoretical point of view, so that the reader has some introductory insight in this field. First, history concerning the measures and crucial ideas around them is discussed. Afterwards, methods of calculation most often used in connection with health expectancies are presented. The chapter is wrapped up with an overview of some of the most popular measures of health expectancy, and lists some institutions that compute and publish these indicators on a regular basis.

2.1 Health expectancy: history of the concept

The concept of health expectancies is nowadays already well established, since its history stretches back to 1960s. Sanders (1964) is considered to be the first one to come up with the concept of healthy life span as an addition to total life span. In his paper, questions are raised about the relationship among morbidity, mortality and level of health care. Ultimately, suggestions to create a procedure making it possible for researchers and statisticians to measure efficiency of health care system in terms of productive man-years are expressed (Sanders, 1964). His view of measuring production capacity of individuals was a foundation that provided a sound basis for ideas and concepts that eventually developed into health expectancy indicators.

The work of Sanders was further developed by Sullivan (1966, 1971). First, he published a study touching on some of the conceptual issues in measuring length of healthy life, and concluded that if a general index is to be created, it should comprise a mortality measure, as well as some kind of morbidity measure. In order to measure morbidity, use of level of disability in population was suggested (Sullivan, 1966). Later, he proposed a framework for calculation of healthy length of life (Sullivan, 1971), and is considered to be the first one to come up with a summary measure in this field. Using age- and sex-specific disability rates, he modified the standard life table computation process in order to obtain life expectancy free of disability, and expectancy of disability (Sullivan, 1971). This is in later literature referred to as the Sullivan method, which, until now, remains the most popular method for health expectancy computations. Details concerning the Sullivan method are described in more detail further on in the text.

Health expectancies gained popularity especially at the end of 1980s (Mathers, 2002), and in 1989, Network on Health Expectancy (REVES, Réseau Espérance de Vie en Santé) was established in France (INED, 2019). This network was established with the aim of promoting use of health expectancy and harmonizing calculation methods in order to attain cross-national comparability of the indicators (Bone, 1992; INED, 2019). As of today, REVES consists of approximately 150 researchers from 30 countries, who represent a list of professional disciplines, among others demography, public health, sociology, medicine or gerontology (INED, 2019).

As the discussion about morbidity, mortality and healthy life span emerged and continued, three competing theories on their relationship and future development arose – compression of morbidity, expansion of morbidity, and dynamic equilibrium. The theory of compression of morbidity was introduced by Fries (1980) and says that ever higher proportion of life is lived without disease. It builds on a premise that there is a natural limit to the length of human life span and that it is possible to compress morbidity and disability to the years immediately preceding death (Fries, 1980).

In the theory of expansion of morbidity, the opposite is believed to be true. Gruenberg (1977), as a proponent of morbidity expansion, argues that the years of life spent with disease have increased because of the capability to postpone fatal complications. Incidence of chronic conditions and their severity do not change over time, in spite of medical successes. In his view, the net effect of technical advances in medicine was actually to make people's health worse (Gruenberg, 1977).

Manton (1982) reflects on both of these theories and expresses a skeptical attitude to the thought of Fries that there is a biological or social constraint to life expectancy. However, he does not agree with all of the conclusions made by Gruenberg, either. Instead, concept of a dynamic equilibrium, where mortality and morbidity are linked in a more sophisticated manner, is established. Dynamic equilibrium states that advances in life expectancy tend to be caused neither by reducing the incidence of diseases, nor by eliminating lethal events in the course of the disease, but by lowering the rate of progression of the disease (Manton, 1982). This implies that prevalence of chronic diseases rises (which would be in line with morbidity expansion theory), but the disease severity is reduced. Proportion of life lived with disability thus does not change significantly if this theory holds.

Naturally, health expectancies are used to evaluate validity of the three theories – Jagger and Robine (2011) put it so that these indicators were even developed for this purpose. In this context, concepts of compression and expansion of morbidity were enriched in order to form a clear relationship between them and health expectancy. Absolute and relative compression and expansion, respectively, were defined. With absolute compression of morbidity, the number of years spent with ill-health decreases, whereas with absolute expansion of morbidity, this number increases. In contrary, relative compression implies decreasing proportion of life spent with ill-health, relative expansion means that this proportion increases (Howse, 2006). Under the assumption of a steady increase in total life expectancy, absolute compression results always also in relative compression. Absolute expansion can be connected both with relative compression and relative expansion.

There is a number of studies assessing whether compression, expansion or dynamic equilibrium is more likely to hold (e.g. Jagger and Robine, 2011; Salomon et al., 2012). However, Howse (2006) points out that the three competing theories are not absolutely mutually exclusive, they are mutually exclusive only when discussing the main driving force behind the postponement of death; underlying factors included in the theories may to some extent operate simultaneously, implying that each of the theories partly holds.

2.2 Health expectancies: methods of computation

By definition, health expectancies are derived from life expectancies, so it almost goes without saying that methods of health expectancy calculation are to an extent based on standard life expectancy calculations, i.e. life table methods. These are the three most common ways to calculate health expectancies: Sullivan method (also called cross-sectional method), multistate method and multiple-decrement method (Mathers, 2002).

The Sullivan method was originally presented by Sullivan (1971) and was already mentioned earlier in the text. Method is based on standard life table and prevalence of health state which is of interest, formulas provided below are based on Mathers (2002, p. 187). Let the variables necessary for computation be introduced in the following way:

l_x	number of survivors at exact age x
L_x	number of person-years lived at age group x to $x+n$
$prev_x$	prevalence of health state D at age group x to $x+n$

Number of survivors and number of person-years lived is taken from a life table with age interval of n years. In order to calculate health expectancy using Sullivan method, person-years lived at age group x to $x+n$ in health state D (YD_x) and outside of health state D (YWD_x) are computed as (Mathers, 2002, p. 187):

$$YD_x = L_x \times prev_x$$
$$YWD_x = L_x \times (1 - prev_x)$$

In a second step, life expectancy in health state D (LED_x) and outside of state D ($LEWD_x$) are calculated as the sum of person-years lived in given health state at or above age x divided by number of survivors at exact age x, ω being the last open-ended age interval (Mathers, 2002, p. 187):

$$LED_x = \frac{\sum_{i=x}^{\omega} YD_i}{l_x}$$

$$LEWD_x = \frac{\sum_{i=x}^{\omega} YWD_i}{l_x}$$

Prevalence has to be disaggregated by sex and age group with age intervals corresponding to those used in life table. Method can be used with various definitions of health states. Original Sullivan's proposal included two states – with and without disability – but polychotomous scale of health states can arbitrarily be used as long as the sum of prevalences across all health states is equal to one (Jagger and Robine, 2011). Advantages of this method are its simplicity and relative ease of data collection, since apart from life tables, period prevalence of health state of interest is the only input required. On the other hand, there are concerns about statistical quality of Sullivan method, but they have been addressed in a handful of studies without serious implications for practical use (Mathers and Robine, 1997; Imai and Soneji, 2007).

Multistate method first appeared in connection with health expectancy in 1989 (Rogers et al., 1989). This method models transitions in and out of multiple health states, as well as mortality. The idea is based on the fact that people can recover from a disease. Ill-health is not an absorbing state, and therefore any person can go from ill-health back to good health. Longitudinal data are needed in this method, and this requirement is likely the reason for method‘s limited usage in practice. Nevertheless, provided data of appropriate quality are available, benefits of multistate method are the possibility to compute health expectancy for person in specified health state at a given age or the possibility to take into account differing mortality rates in different health states (Mathers, 2002). Moreover, multistate methods are well suited for comparisons between subpopulations (Jagger and Robine, 2011).

Multiple-decrement method is a special case of multistate method, where there is no possibility of transition from state of ill-health back to health (Mathers, 2002). An early application of this method was performed by Katz et al. (1983) on active life expectancy, but its main area of use are irreversible health states, such as dementia. For calculation of dementia-free life expectancy is this method especially suitable (Jagger and Robine, 2011).

2.3 Health expectancies: classification

Summary measures of population health can be divided in two areas depending on the logic of their construction. The first area are health expectancies (these are in focus in this text), the second one are health gaps. Health gaps express a difference between actual health status of a population and a hypothetical state of health defined by a norm or a goal for population health (Mathers, 2002). However, health gaps will not be discussed further on.

The term health expectancy is used as a general term for indicators that express average time in years expected to live in certain state of health (Mathers, 2002). This description implies there is no universal approach towards health expectancy computation, or as Jagger and Robine (2011, p. 552) put it, „there are as many health expectancies as health measures“. Health expectancies thus encompass a variety of approaches, out of which some are presented in this section.

Mathers (2002), building on Mathers et al. (1994), categorized health expectancies in two classes. One class consisted of indicators using dichotomous or polychotomous weights of health states, i.e. that there is a number of health states, for which, individually, life expectancy is constructed. Examples of this class include disability-free life expectancy (DFLE) and life expectancy in good self-perceived health, both of which will be described in more detail further on in the text.

Second class of health expectancies includes indicators that use a finite exhaustive set of health states, and assign them with values according to their severity. Based on these valuations, expectation of equivalent of years spent in good health is computed. Disability-adjusted life expectancy (DALE), sometimes referred to as health-adjusted life expectancy (HALE), belongs to this class.

As mentioned above, disability-free life expectancy is an example of the first class of health expectancies and is, among health expectancy indicators, one of the most commonly used. In order to obtain DFLE, proportion of population living with disability (disaggregated by sex and age groups) is needed, and the most common way is to use the Sullivan method. DFLE can be interpreted as an average number of years one can expect to live without limitations in everyday activities or disability (Mathers et al., 1994).

Disability-free life expectancy plays an important role in the statistics of European Union (EU), since it was chosen as one of the European structural indicators. In the EU, indicator is called healthy life years (HLY). Eurostat’s computation of HLY is based on self-perceived long-standing activity limitation data, which are acquired from EU statistics on income and living conditions (EU-SILC) survey. Apart from that, Eurostat computes healthy life expectancy based on self-perceived health. This indicator combines mortality data with the proportion of population claiming to be in good state of health. As with HLY, the data on health come from EU-SILC (Eurostat, 2012; Eurostat, 2014a).

A representant of the second class of health expectancies as classified by Mathers (2002) is health-adjusted life expectancy (HALE). This indicator is computed on global scale within the Global Burden of Disease (GBD) project, which generally aims to provide data on health status for all of the world’s countries. HALE (in context of GBD also called only *healthy life expectancy*), in order to be calculated, needs a set of health states that are assessed according to their severity, as described earlier. Researchers within GBD used multiple sources of health data in order to make this assessment, among others survey data, findings from relevant literature, data from hospitals or disease registries (Salomon et al., 2012; Vos et al., 2012). This multi-dimensional approach was developed as a consequence of very limited comparability of data on self-reported health across countries, which is mainly caused by differences in expectations and norms for health (Mathers et al., 2001). Once the average health in each age group was estimated, it was further incorporated in the Sullivan method, resulting in HALE estimates (Salomon et al., 2012). The GBD project, originally established in 1990 under World Health Organization (WHO), continues its work and aims to present up-to date estimates of ever higher quality on health expectancy indicators (Murray et al., 2015; Kassebaum et al., 2016).

Even though the concepts of DFLE and HALE are the most usual ways to express healthy life expectancy in practice and in scientific work, there are other concepts worth mentioning, since, as was already pointed out, there is theoretically unlimited number of health expectancies.

One example is active life expectancy, which aims to express the number of years remaining to the point of loss of independence in activities of daily living (Katz et al., 1983). This concept is not very dissimilar to DFLE, but the definitions differ somewhat.

It is also fairly common to calculate life expectancies with absence of certain diseases, for example dementia-free life expectancy (Ritchie et al., 1994). Or in a broader sense, life expectancy without chronic diseases, taking into account either all or a set of such diseases, can be met in research (Perenboom et al., 2005). Life expectancy without chronic morbidity can be calculated on European level from data in EU-SILC, since the survey contains a question on presence of a chronical illness (Robine and Jagger, 2017), but the calculated indicator is not accessible in Eurostat database. Table 1 summarizes health expectancies discussed in this section.

Table 1 – Overview of selected health expectancies and underlying measures

Health expectancy	**Underlying health measure**
First class of health expectancies according to Mathers (2002)	
Disability-free life expectancy (healthy life years)	presence of disability (limitation in activities of daily living)
Life expectancy according to self-perceived health	subjective perceivement of health status
Life expectancy without chronic morbidity	presence of a chronic disease
Active life expectancy	independence in activities of daily living
Second class of health expectancies according to Mathers (2002)	
Health(disability)-adjusted life expectancy	complex formal valuation of health states in each age group

Source: author's overview based on Mathers (2002)

Chapter 3

Literature review

There is extensive literature on the linkages between population health and macro-economic, educational, and other factors. In this chapter, review of such literature is provided in order to summarize the work that has already been done in this field. At the end of the chapter, hypotheses for this work are stated.

It is necessary to emphasize that most studies evaluating population health consider more than one area of explanatory variables, since the effect of individual variables is hard to distinguish without correcting for some co-variables (studies examining, for instance, association of level of poverty and population health, will most likely include indicators of performance of labour market or variables taking into account health system in a given country). However, this chapter analyzes the groups of explanatory factors individually, so that also the mechanism linking the fields can be discussed.

3.1 Macro-economic variables and health expectancies

The gross domestic product (GDP) per capita, or another aggregate expression of wealth of a country, is very often used in analyses of all kinds because of its information value and simplicity to understand even for the general public (Subramanian et al., 2002; Cutler et al., 2006). As such, it also has a number of drawbacks, which are nowadays well described (e.g. in Mankiw, 2012).

The positive linkage between GDP per capita and health is considered to be well proven (Bloom and Canning, 2000), and in the European context was documented e.g. by Jagger et al. (2008), who analyzed healthy life years as an indicator of population health. The pathway in this case is that higher GDP improves accessibility of goods and services that positively influence health of individuals (Jagger et al., 2008). The relationship, nevertheless, is not straightforward.

Apart from the GDP itself, attention needs to be focused on the level of equality of wealth distribution within a country, which also tends to be an important determinant of the health of population (Cutler et al., 2006; Jagger et al., 2008). Even in the developed countries of the world, there exists a gradient in population health, which puts the poor in a position of disadvantage (Subramanian et al., 2002; Cutler et al., 2006). Indeed, according to the study of Matthews et al. (2005), the subjective measure of adequacy of income (which can be translated into the measure of whether a person finds themselves poor in the context of the society) greatly influenced the onset of disability in his study of people over 75 years of age in the United Kingdom.

As for the mechanisms mediating the effect of income inequality on health measures, the following have been suggested: Firstly, the material deprivation leads to worsened access to health care because the services are consumed primarily by those with appropriate means (Subramanian et al., 2002; Cutler et al., 2006). Another mechanism discussed is that the lack of means is connected with lower social capital or social exclusion, which further inhibits the access to health care (e.g. Kawachi et al., 1997). Lastly, Wilkinson et al. (1998) propose that there is a direct psychosocial pathway – the hopelessness, lack of respect or loss of control over one's own life affect individual health directly.

It is of importance to mention, that the relationship between income and health does not work unidirectionally, but rather that these two reinforce themselves mutually (Bloom and Canning, 2000; Cutler et al., 2006). There are several mechanisms that are believed to mediate the influence of health on income. Firstly, healthier population has higher work productivity, which in turn increases income. Secondly, better health incentivizes investments in education because of greater chance to enjoy the product of such investment – and more educated society generates higher productivity and higher income. And thirdly, longer life reinforces the need for savings for old age, and if this saved capital is invested, workers have access to more capital and national income thus rises (Bloom and Canning, 2000). It can be seen even in this short list, that the explanations involve education as a mediating factor. From that, it is clear that the fields that are discussed separately in this chapter, operate simultaneously in practice.

Part of the connection between income and health can be explained by the amount or proportion of expenditures on health care, where the effect is expected to be positive. Higher governmental health care spending should provide better supply of health care, expressed for example by the number of physicians, number of visits to a doctor or the overall accessibility of health care (Crémieux et al., 1999). Anderson and Frogner (2008), however, point out that the utilization of the spending is crucial for determining the health outcomes, and show this on the example of the United States, where very high spending is not justified by appropriate values of health indicators.

In practice, this relationship was studied extensively, and although some researchers found evidence of positive impact of health care spending on health outcomes in specific contexts (Crémieux et al., 1999; Gupta et al., 2002), caution is often suggested in relation to generalization of results in a cross-national setting (Anderson and Frogner, 2008; Jagger et al., 2008), implying that the relationship is far from being clear and universal, and that specific, more detailed characteristics of design of health care systems should be accounted for.

3.2 Unemployment and health expectancies

In a number of studies, researchers made attempts to link health status and unemployment, including some case studies within Europe. Jagger et al. (2008) found some evidence on the link between HLY and long-term unemployment rate, but in their study of EU countries, results were not entirely conclusive. On the other hand, unemployment was found to account for a significant portion of inter-regional variance in disability-free life expectancy in Spain (Gutierrez-Fisac et al., 2000). On an individual level, evidence was found within the EU, that unemployed people are more likely to report poor subjective health (Alavinia and Burdorf, 2008). Although such documentations of correspondence between unemployment and health can be found, other studies suggest more complicated relationship between these two (Béland et al., 2002; Roelfs et al., 2011).

As with income, unemployment is also interlinked with health in the opposite direction, meaning that poor health or disability can limit one's involvement in the job market. Laditka and Laditka (2016) have documented such a situation in their longitudinal study of disability in the United States.

Pathways linking the two processes are multiple – association of unemployment has been shown with obesity, physical inactivity and stress (Kalousova and Burgard, 2014; Alavinia and Burdorf, 2008), which influence general health negatively. Apart from that, the economic hardship stemming from loss of employment is believed to increase the risk of adopting hazardous behaviour, like alcohol or tobacco consumption; another possible outcome of unemployment is a decrease in social capital due to loss of social connections and habits (Kalousova and Burgard, 2014). The influence that a decrease in social capital can have on health was already mentioned in this text.

3.3 Education and health expectancies

Yet another field, which is connected with health outcomes and whose impact on health has been studied, is education. Conti et al. (2010, p. 234) write that „a positive correlation between health and schooling is one of the most well established findings in the social sciences", and this situation is acknowledged by many other authors as well. However, this does not necessarily imply that the relationship is straightforward, easy to uncover or perfectly described.

As with phenomena that were described previously in this chapter, there is not a unique way through which the association between education and health works. Three attitudes, in a broad sense, are expressed in literature. The first one states that there is a causal relationship that leads from education to health. The second one stresses the opposite – that the causal relationship runs from health status to level of education. Finally, a third attitude argues that there is a „third variable" that affects both schooling and health in the same direction (Grossman, 1976; Cutler and Lleras-Muney, 2006). It is unlikely that only one of these three holds, it is rather believed that each of these mechanisms can explain a part of the association (Grossman, 1976).

Bearing in mind these multiple possible explanations of the relationship, various researchers have found the direction from education towards health to be more significant than the other ways. Grossman (1976), Ross and Wu (1995), Cutler and Lleras-Muney (2006) and Conti et al. (2010) have all found such evidence in their studies analyzing the United States and the United Kingdom, even though their time periods of interest, specific data sources and methods used differed. One possible explanation is brought by Ross and Wu (1995), who justify this direction of causality by saying that education is usually completed in the first three decades of life, when disability and illness is low, and that it is therefore unlikely that poor health prevents a significant number of people from finishing school.

The situation in the EU seems to be less clear, especially when running a cross-national analysis. On the national level, education was found to be positively linked to health only in the countries that were members of the EU prior to the Eastern enlargement in 2004. In the ten countries that joined EU in 2004, this relationship did not hold seemingly because of the generally high level of education in these countries, which was not accompanied by respective advances in population health (Jagger et al., 2008).

In general, there are many pathways that help in explaining the effect education has on health, and these pathways are complements to each other. The first one operates via raising productivity – education leads to higher productivity of the individual or of the society as a whole, which in turn increases income and thus also possibilities with regard to health care accessibility (Grossman, 1976; Cutler and Lleras-Muney, 2006). Higher education makes it also more likely for an individual to be able to work in a safe environment (Cutler and Lleras-Muney, 2006) and it has been found that less educated people are more prone to engaging in risky behaviour, e.g. alcohol and tobacco consumption (Hunt-McCool and Bishop, 1998). Possibly as a consequence of this, those with higher education value their health more, and the incentives to invest in their own health (regular check-ups, good health insurance etc.) are therefore higher for these people, compared to those with lower education (Cutler and Lleras-Muney, 2006). Educated people are also believed to absorb any new information about healthy lifestyles or promotions in treatment more quickly, which is another reason for their health to be better (Cutler et al., 2006). And finally, higher education leads to a broader social network, which can be of support and assistance in case of any medical issues (Cutler and Lleras-Muney, 2006).

Attention is also given to the opposite direction of the association, but the research seems not to be as extensive. The performance at school was linked to indicators of health either of the child itself (for example expressed as weight at birth) or of its parents, and some positive linkage was found (Conti et al., 2010). Not even here, however, is the link straightforward – Wolfe (1985) has found differing impacts on child's performance at school, based on what their medical condition specifically was. Some of the conditions were found to have large impact, but others seemed not to be of importance for school results (Wolfe, 1985).

3.4 Summary and formulation of hypotheses

This chapter summarized some of the scientific work that has been done by now in the subject of explaining population health with characteristics of macro-economy, labour market and education system. The amount of research that has studied such relationships is sizeable, and the aim of this chapter was not to comprise all of it. However, substantial issues which have to be considered while performing an analysis in this field were addressed and they were documented on a number of examples.

Considering the large amount of research, it is quite surprising that – even in recent history – it was not health expectancies that were most often used as measures of overall population health in the studies. Instead, indicators used consisted of infant mortality, total life expectancy, and sometimes even crude death rates, all of which have limitations, especially in developed countries. Results from studies which took into account such imperfect measures must therefore be interpreted with caution. This is not to say that studies analyzing health expectancies are nonexistent, but that their occurrence is surprisingly low. Using health expectancies can presumably add more insight in these phenomena simply because of their nature of being measures that handily combine morbidity and mortality.

Based on the review provided in this chapter, hypotheses which are tested in the analytical part of this study are presented and commented on below. Hypotheses are set in line with the overall goals of the work introduced in the first chapter, and testing for their validity will contribute to achieving the goals.

- **Hypothesis 1**: *Gross domestic product per capita is positively associated with health expectancy.*

  This is a basic relationship which, based on the literature studied and on beliefs of a number of researchers expressed in literature, is expected to hold.

- **Hypothesis 2**: *Income inequality is negatively associated with health expectancy.*

  This is a macro-economic topic which scientists recently appear to be very interested in. The assumption in this case is that, after controlling for GDP per capita, population of countries with higher income inequality experiences worse health.

- **Hypothesis 3**: *Education has positive impact on health expectancy.*

  As was mentioned in section 3.3, education is generally believed to impact health positively. However, questions were raised as for whether the relationship holds even in specific European setting (Jagger et al., 2008).

- **Hypothesis 4**: *Unemployment rates are negatively associated with health expectancy.*

  This hypothesis is supposed to be tricky, since there is no consensus in the literature about this relationship. Hypothesis formulation is thus based on some studies which were undertaken in European setting (Gutierrez-Fisac et al., 2000; Jagger et al., 2008) and which appeared to support this relationship, at least to some extent. Time lags, as well as some other characteristics of the job market as a whole, will possibly play their role in assessing this association.

Chapter 4

Data description

Since the geographical scope of this thesis are European countries, Eurostat database was chosen to be the source of data for the analysis. The database provides data for all member states of the EU, as well as some other European countries, in a broad list of fields. The aim of Eurostat is to harmonize data which are collected on the level of individual countries, so that cross-national analyses can be undertaken without having to correct for differences in definitions. Nevertheless, some differences in how the data are collected still exist, and they are, if applicable, commented on in the rest of this chapter, in respective sections discussing the dataset in more detail.

Creation of the dataset for this study stems from the hypotheses formulated in previous chapter; the set is constructed in such a way that validity of the hypotheses can be evaluated. Apart from data on healthy life years and healthy life expectancy based on self-perceived health, areas included are GDP per capita, share of GDP spent on health care, unemployment and long-term unemployment, education attainment, income inequality, risk of poverty or social exclusion and also life expectancy.

Data were collected for all countries of the EU (including United Kingdom), and for Iceland, Switzerland and Norway, thus for 31 countries in total. As for the timespan, 2005 is the first year when most of the comparable data are available for countries in Central and Eastern Europe that joined EU in 2004, which is the reason for choosing 2005 as the beginning of the period under study. After 2005, data are collected for each year until 2017, which is at the time of data collection the latest year with available data on health expectancies.

4.1 Health expectancy and life expectancy data

Healthy life years, healthy life expectancy (HLE) based on self-perceived health and total life expectancy are all collected at birth, at age 50 and at age 65, separately for men and women. Life tables, and corresponding life expectancy, are computed by Eurostat from raw mortality data supplied by member states (Eurostat, 2017a). HLY and HLE based on self-perceived health are calculated using mortality data and combining them with health and disability data which result from the EU-SILC survey (Eurostat, 2012; Eurostat, 2014a). Sullivan method, as described in Chapter 2.2 of this text, is used for calculations.

In case of healthy life years, the specific question in the survey providing data for computation is „For at least the past six months, to what extent have you been limited because of a health problem in activities people usually do? Would you say you have been:" with possible answers *severely limited*, *limited but not severely* and *not limited at all* (Eurostat, 2014a). Health measure resulting from this question is referred to as Global Activity Limitation Index (GALI; Jagger et al., 2010). Proportion of population in healthy condition is based on answer *not limited at all*, whereas proportion in unhealthy condition is based on the other two answers (Eurostat, 2014a).

With healthy life expectancy based on self-perceived health, the relevant question in the survey is „How is your health in general?" with possible answers *very good*, *good*, *fair*, *bad* and *very bad*. Choice of very good, good or fair health is evaluated as good health for purposes of computation (Eurostat, 2012). Eurostat classifies answers resulting from this question in only two categories, but if needed, data from EU-SILC make it possible to compute health expectancies for all five categories of health that can be chosen in the answer.

The EU-SILC has an important drawback caused by design of the survey – it comprises only private households. Population living in institutions is therefore systematically omitted, and while computing HLY, private households are believed by Eurostat to be sufficiently representative of the total population (Eurostat, 2014a). Cambois et al. (2008) compared health expectancies based on various surveys in France and concluded that not including instutional population does not affect conclusions. In other studies evaluating health expectancies in Europe, this drawback was either acknowledged without implications for results (Jagger et al., 2008) or neglected (Fouweather et al., 2015).

Another issue that can hamper the cross-national comparability of healthy life years is the difference in wording of the standard GALI question arising as a consequence of translation into individual languages. Eurostat is tracking all the differences and makes them partly available[1] (within database metadata, changes are tracked only until 2012). Tracking of changes performed on the GALI question in recent years (after 2012) is limited, possible explanation for which can be lower capacity within Eurostat allocated to HLY and GALI (as reported by Bogaert et al., 2018).

Although cross-national comparability of the question increased over time, especially after a wave of harmonization in 2008, the efforts still continue as complete harmonization is not yet reached (Bogaert et al., 2018). The potential error stemming from this issue has been studied by Jagger et al. (2010), who in spite of the differences found GALI to be satisfactorily comparable among most of the countries. There are also cultural differences between countries that may influence HLY even though the definition of GALI is the same, but this is an issue that can be tackled in only a very limited way (Jagger et al., 2010).

---

[1] The document tracking differences in wording of the GALI question between countries up to 2012 is available at https://ec.europa.eu/eurostat/cache/metadata/Annexes/hlth_hlye_esms_an2.pdf as an annex to HLY metadata.

Some changes in the GALI question, however, were so serious that they make comparisons over time in certain countries practically impossible. Such a change occurred in Germany in 2015 – until then, wording of the question was significantly different from the standardized English wording (Eurostat, 2019d). Other breaks in time series are discussed and illustrated with figures in the following chapter.

Bearing all these limitations in mind, this thesis uses HLY and HLE based on self-reported health exactly as they are reported in Eurostat database. In analyses further on, HLY data are to some extent put aside so that results are not biased because of the differences in GALI question. However, no adjustments to the values are made.

There are some missing values for these variables, but these appear not to be systematic. For example, in 2010, all HLY data are missing for Italy, in 2011 the same holds for Norway, and in following years, this happens also with Sweden and Finland. For the countries that joined EU after 2004 (Romania, Bulgaria and Croatia), availability of data is also limited in the beginning of the period analyzed. Moreover, the most recent year when HLE based on self-reported health is available is 2016, compared to 2017 for all other variables in the dataset.

4.2 Gross domestic product, income and inequality data

In the field of wealth of a country and inequality of wealth distribution, the following indicators were chosen: GDP per capita in Euros, GDP per capita in purchasing power standard, Gini coefficient of income distribution and at-risk-of-poverty or social exclusion rate.

Gross domestic product is reported in Euros and in purchasing power standard, where the latter eliminates cross-country differences in price levels. It is standardized using a homogenous basket of goods and services that takes into account price variations between countries (Eurostat, 2019a). GDP per capita is a classical measure broadly used as an expression of a wealth of a country, used in the same way in this study. It is the most complete variable in the dataset, without any missing values within the countries and timespan studied. GDP is also considered to be very well comparable both across countries and over time, thanks to unified definitions (Eurostat, 2019a).

Inequality in income distribution is expressed in terms of Gini coefficient of equivalised disposable income, which in Eurostat is based on EU-SILC, similarly as health expectancies. Disposable income is calculated on the level of households and distributed to individuals afterwards. It includes all income from work, investment and property, and social transfers including old-age pensions, for all household members combined (Eurostat, 2017b). Total income is then equivalised so that differences in household size and composition are taken into account. Eurostat uses a scale where first adult in a household is assigned weight of 1.0, every other person aged 14 or more is assigned weight of 0.5, and the remaining persons are assigned weight of 0.3 (Eurostat, 2017b). This income is then distributed to each member of the household. Data on Gini coefficient of equivalised disposable income are fairly complete in the dataset.

Another variable related to income distribution in society is at-risk-of-poverty or social exclusion rate, which is another indicator arising from EU-SILC. This measure is calculated from the sum of persons, who are either at risk of poverty, suffering from severe material deprivation or living in a household with zero or very low work intensity, relative to total population (Eurostat, 2014b). The risk of poverty threshold is considered to be at 60 % of country's median equivalised income after social transfers; severe material deprivation is defined as a situation, when an individual does not have appropriate means to afford 4 or more items from a list of 9 necessities[2] making it possible to live a decent life in Europe (Eurostat, 2014b).

At-risk-of-poverty or social exclusion rate is available separately for men and women, as well as for both sexes combined. Furthermore, measure is reported taking into account all age groups, but also separately for age group of 65 years and over, with division according to sex available for both age specifications. Coverage of analyzed countries is high, missing values are scarce.

4.3 Education and unemployment data

Within the field of education, percentage of population with low education (lower secondary at most) and with tertiary education are collected and used in the dataset. Within unemployment, variables gathered are unemployment rate and long-term unemployment rate. All of these are common measures that are regularly used as characteristics of labour market and educational structure of population.

Data on educational structure reported by Eurostat are acquired through Labour Force Survey (LFS), another survey which is conducted annually in all countries which are studied in this thesis. Data use the International Standard Classification of Education (ISCED) where level 0 refers to less than primary education and level 8 refers to doctoral degree or equivalent. In the dataset, proportion of population with low education includes levels 0 to 2 – less than primary education, primary education and lower secondary education. Proportion of population with tertiary education includes levels 5 through 8 according to ISCED (Eurostat, 2017c).

Both measures of educational attainment are reported for men, women, and for both sexes combined. Population, from which the proportions are calculated, is all population between 25 and 64 years of age, excluding by definition of LFS those living in institutional or collective households (Eurostat, 2017c). There are almost no missing data in the dataset for these variables.

Unemployment rate and long-term unemployment rate are, same as educational attainment, derived from LFS. Unemployment rate is defined as the share of unemployed people in active

[2] These necessities are: (1) being able to pay rent, mortgage or utility bills, (2) keeping home adequately warm, (3) eating meat, fish or vegetarian equivalent every other day, (4) being able to spend one week's holiday away from home every year, (5) being able to have a telephone, (6) being able to face unexpected expenses, (7) being able to have a colour TV, (8) being able to have a washing machine, (9) being able to have a car

population, where active population are unemployed and employed people combined (Eurostat, 2018). LFS defines an employed person as someone who is 15 years or older and in the reference week of the survey worked at least one hour for pay or profit or family gain, or was temporarily absent from such work (Eurostat, 2019b). On the contrary, an unemployed person is someone who is 15–74 years of age, is not employed according to the definition above, is currently available for work and had been actively seeking for work during four weeks prior to the interview (Eurostat, 2019b). Long-term unemployment rate is the share of those who are unemployed for 12 months or more in active population (Eurostat, 2018).

Data for unemployment cover whole set of countries, apart from Switzerland, for which the data are missing. As with education data, unemployment rate and long-term unemployment rate are both available for men, women and for total population.

4.4 Health care expenditures data

Last set of indicators is related to government-level expenditures on health care. Eurostat provides such data in a detailed classification, where different levels of government and different areas of health care spending are taken into account. However, in the dataset for this thesis, only expenditures of general government on health in total are considered. All of the expenditures are expressed as a percentage of country's GDP.

General government, whose spending on health is reported, consists of four subsectors: Central government, which is the top-level administrative unit in every country; State government, a lower level government existing only in some countries with federal structure (for example Germany, Belgium, Spain); Local government, which includes region-specific administrative units; and Social security funds (European Commission, 2011).

These data, reported by Eurostat, follow the standardized Classification of Functions of Government, and are collected in the framework of European System of National Accounts (Eurostat, 2019c). Thanks to the standardized classification and system of data collection, comparability both across countries and across time is considered to be very high (Eurostat, 2019c). Data are available for all studied countries, without a significant number of missing values.

4.5 Dataset summary

This chapter described in detail the dataset which is used further on in the analysis, with focus on quality and comparability of individual variables within the dataset. Measures of health expectancy (and, secondarily, life expectancy) are considered to be in center of attention, and their inter-country variation is to be explained with help of the other measures discussed. The dataset is constructed in a way that makes it possible for the analysis to yield results, which can in turn be used to evaluate validity of hypotheses stated in Chapter 3 of this text. Table 2 displays an overview of all the variables in the dataset, the complete dataset is available in Electronic Appendix.

As stated in the beginning of this chapter, period covered comprises years 2005 to 2017. Only variable which is in Eurostat database available for less than full time span is healthy life expectancy based on self-reported health – here, the latest year available is 2016.

Since Eurostat uses unified methodology for calculation and reporting of data, all of the variables collected can be considered well comparable across countries. Some caveats were found in this respect, they were commented on in respective sections of this chapter. The most important one seems to be related to one of the central response variables, HLY, and is caused by wording in the standardized GALI question, which is a basis for disability measures. Should any adjustments to the data be made because of this issue, they will be properly reported. But as a base line, data in the remainder of this text are reported and used exactly as they were downloaded from Eurostat.

Table 2 – Summary and characteristics of variables for analysis

	Variable	Division	Availability	Notes
Response variables	HLY	• men / women • ages 0 / 50 / 65	2005–2017	
	HLE – self-reported	• men / women • ages 0 / 50 / 65	2005–2016	
	Life expectancy	• men / women • ages 0 / 50 / 65	2005–2017	
Explanatory variables	GDP in Euros		2005–2017	
	GDP in purchasing power standard		2005–2017	
	Government expenditure on health	• 6 subcategories of health	2005–2017	
	Gini coefficient of income distribution		2005–2017	
	At-risk-of-poverty or social exclusion rate	• men / women / total • all ages / ages 65+	2005–2017	
	Unemployment rate	• men / women / total	2005–2017	Missing: Switzerland
	Long-term unemloyment rate	• men / women / total	2005–2017	Missing: Switzerland
	Proportion of population with low education	• men / women / total	2005–2017	
	Proportion of population with tertiary education	• men / women / total	2005–2017	

Source: author's overview

Chapter 5

Trends in health expectancy, life expectancy and explanatory factors in Europe in recent history

Until this point in the text, health expectancies were only discussed theoretically in relation to techniques of calculation or in relation to collection of data needed for calculations. This chapter aims to turn to some more practical issues by providing a brief overview of how values of health expectancy developed in European countries in the past years. The dataset described in previous chapter serves as a primary source for this overview, which is therefore structured as a basic exploratory statistical analysis of the data. Further background and possible trends or contexts are also discussed in addition to pure description.

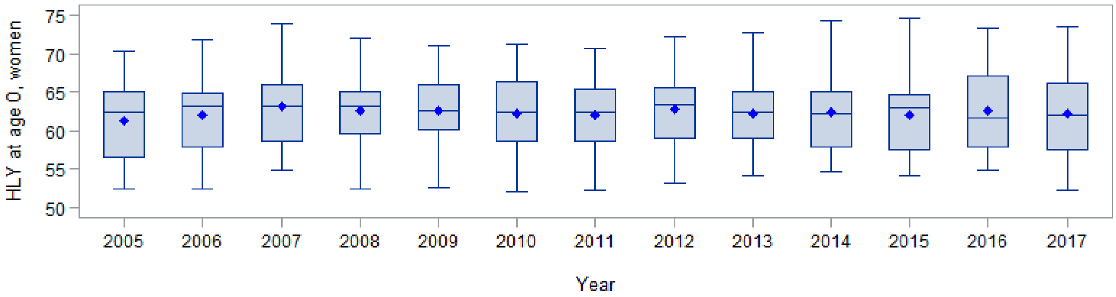
In the second part of this chapter, focus is turned to factors that are used in subsequent chapters for explaining the variation in health expectancy – hence the term explanatory when describing them as a whole (even though strictly speaking, they become explanatory factors only once the statistical model is formulated). A brief description of their trends is provided so that reader gets acquainted with the data more closely before turning to analysis itself.

5.1 Trends in health expectancy in recent European history

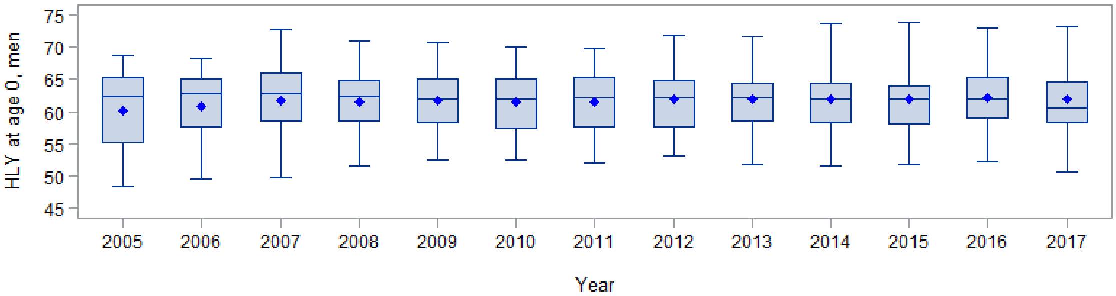
5.1.1 Assessing Healthy life years

Judging visually on the basis of summary statistics, healthy life years did not follow any clear trend between 2005 and 2017 when taking into account all of the studied countries together. This statement holds no matter the age and sex for which HLY is calculated, as is illustrated with a series of boxplots in Figure 1. There is a sign of an upward trend between 2005 and 2007, but thereafter, summary measures are fluctuating and do not allow to form any conclusion. The above said can be shown on median and mean values of HLY in 2017 – for none of the six age-sex specifications is the mean or median HLY in 2017 the highest in the time span. This being stated, it is important to add that values in 2016 tend to be slightly higher than in the years before, although not for all specifications.

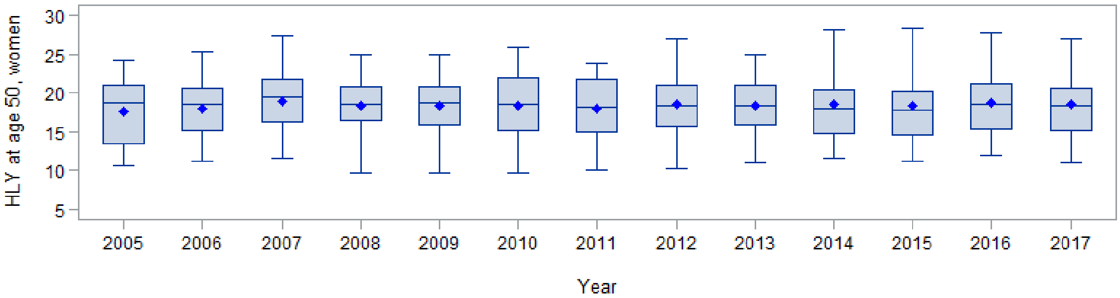
Figure 1 – Summary statistics of HLY at ages 0, 50 and 65, men, women, 2005–2017



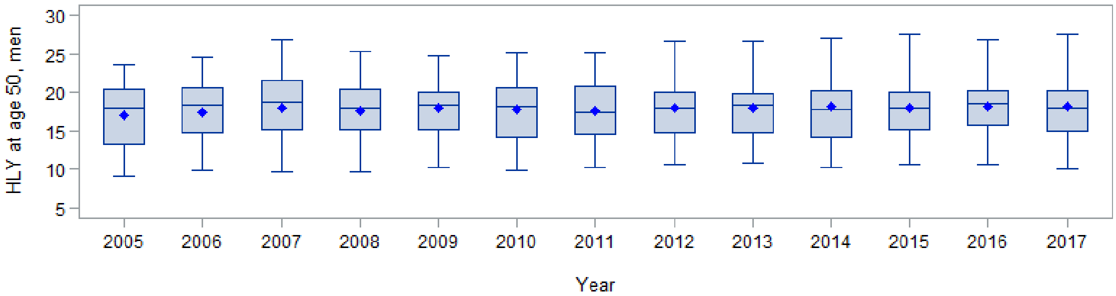
Min	52.4	52.5	54.8	52.5	52.6	52.0	52.3	53.1	54.2	54.6	54.1	54.9	52.2
Mean	61.2	61.9	63.1	62.6	62.6	62.3	62.0	62.8	62.2	62.4	62.0	62.6	62.2
Max	70.4	71.9	73.9	72.1	71.0	71.3	70.7	72.2	72.7	74.3	74.6	73.3	73.6



Min	48.3	49.6	49.8	51.6	52.4	52.4	52.1	53.1	51.7	51.5	51.8	52.3	50.6
Mean	60.2	60.9	61.7	61.5	61.8	61.4	61.5	62.0	61.9	62.0	61.9	62.2	61.9
Max	68.6	68.3	72.8	71.0	70.7	70.1	69.9	71.9	71.7	73.6	74.0	73.0	73.2

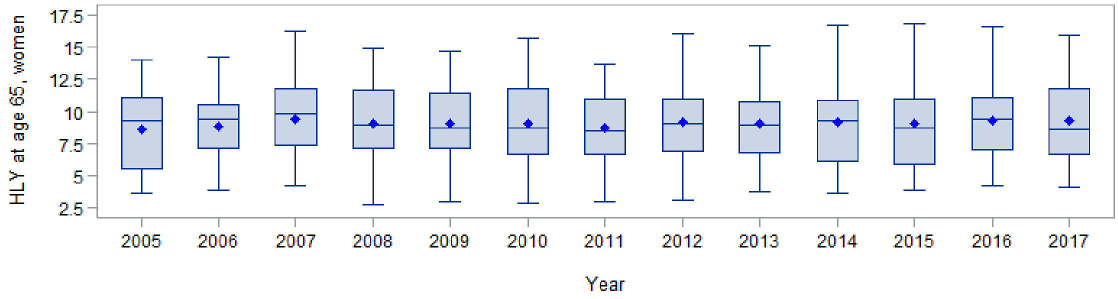


Min	10.6	11.2	11.5	9.6	9.7	9.7	10.0	10.2	11.1	11.5	11.2	12.0	11.0
Mean	17.6	18.0	18.9	18.4	18.4	18.4	18.0	18.6	18.4	18.5	18.3	18.7	18.6
Max	24.1	25.3	27.4	24.9	24.9	25.9	23.9	27.0	25.0	28.2	28.4	27.7	27.1

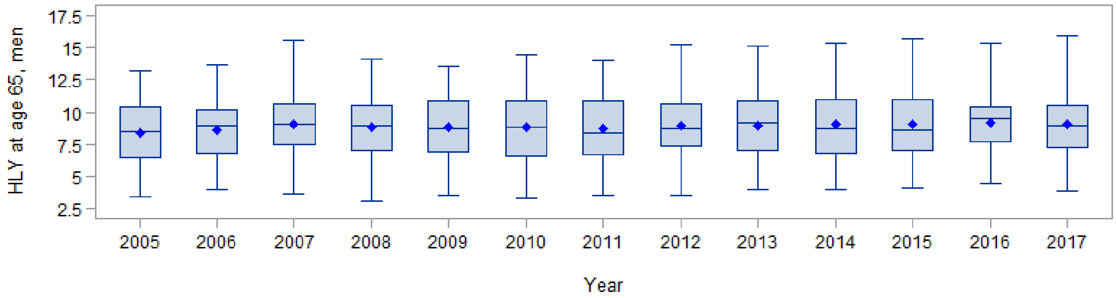


Min	9.2	9.8	9.7	9.6	10.2	9.9	10.2	10.6	10.9	10.3	10.6	10.7	10.0
Mean	17.0	17.4	18.0	17.7	17.9	17.7	17.6	18.0	18.0	18.1	18.1	18.2	18.1
Max	23.7	24.5	26.9	25.4	24.8	25.2	25.1	26.6	26.7	27.1	27.6	26.8	27.5

Figure 1 (cont.) – Summary statistics of HLY at ages 0, 50 and 65, men, women, 2005–2017



Min	3.6	3.8	4.2	2.7	2.9	2.8	2.9	3.1	3.7	3.6	3.8	4.2	4.1
Mean	8.5	8.9	9.4	9.1	9.0	9.0	8.7	9.2	9.1	9.2	9.0	9.3	9.2
Max	14.0	14.2	16.3	14.9	14.7	15.7	13.7	16.0	15.1	16.7	16.8	16.6	15.9



Min	3.4	4.0	3.6	3.0	3.5	3.3	3.5	3.5	4.0	4.0	4.1	4.4	3.8
Mean	8.3	8.6	9.0	8.8	8.9	8.8	8.7	9.0	9.0	9.1	9.1	9.2	9.0
Max	13.2	13.7	15.6	14.1	13.6	14.5	14.0	15.3	15.1	15.4	15.7	15.4	15.9

Source: Eurostat, author's own calculations

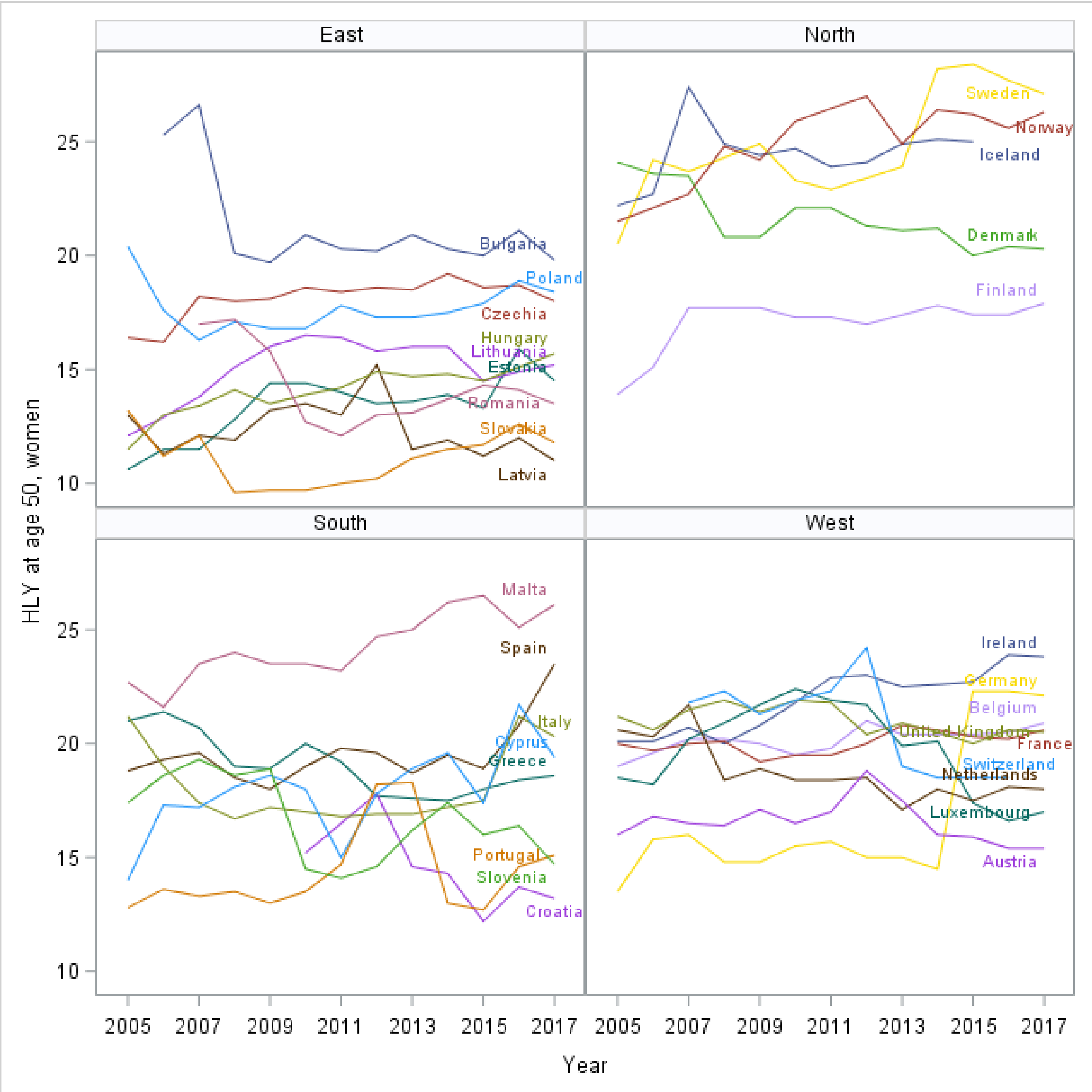
Note: Min, Mean and Max refer to minimum value, arithmetic mean and maximum value of the given HLY specification in given years. Processed by SAS 9.4.

Mean value of HLY at birth increased for women from 61.2 years to 63.1 years between 2005 and 2007, and finally reached 62.2 years in 2017. For men, values in corresponding years were 60.2, 61.7 and 61.9 years, respectively – a similar increase between 2005 and 2007 was observed, but it was not followed by a downward trend aftewards. At age 50, HLY for women first grew from 17.6 to 18.9 years, and afterwards fluctuated and reached 18.6 years in 2017. For men at the same age, mean HLY between 2005 and 2017 increased from 17.0 to 18.1 years (Figure 1).

Results document a fact that was already observed by a number of researchers (e.g. Robine and Cambois, 2013; Van Oyen et al., 2013) – women experience slightly higher values of HLY, but their advantage to men is relatively lower than in life expectancy. Women thus tend to live a higher proportion of life with activity limitation.

It is also worth mentioning the range of values in individual countries – Europe still experiences significant differences between countries. In 2005, HLY at birth ranged from 52.4 (Estonia) to 70.4 (Malta) years for women and from 48.3 (Estonia) to 68.6 (Malta) years for men. The range was even greater in 2017 – from 52.2 (Latvia) to 73.6 (Malta) years for women, and from 50.6 (Latvia) to 73.2 (Sweden) for men (Figure 1). Situation with HLY at age 50 or 65 is, as for the width of the range of values, even more pronounced. For women at age 65, the lowest values of HLY were 3.6 (Estonia) years in 2005 and 4.1 (Slovakia) years in 2017. The highest values, on the contrary, were 14.0 (Denmark) and 15.9 (Norway) years, respectively. Judging by these results, the efforts to narrow the gap among European countries are not very fruitful so far.

Figure 2 – Healthy life years at age 50 in European countries, women, 2005–2017



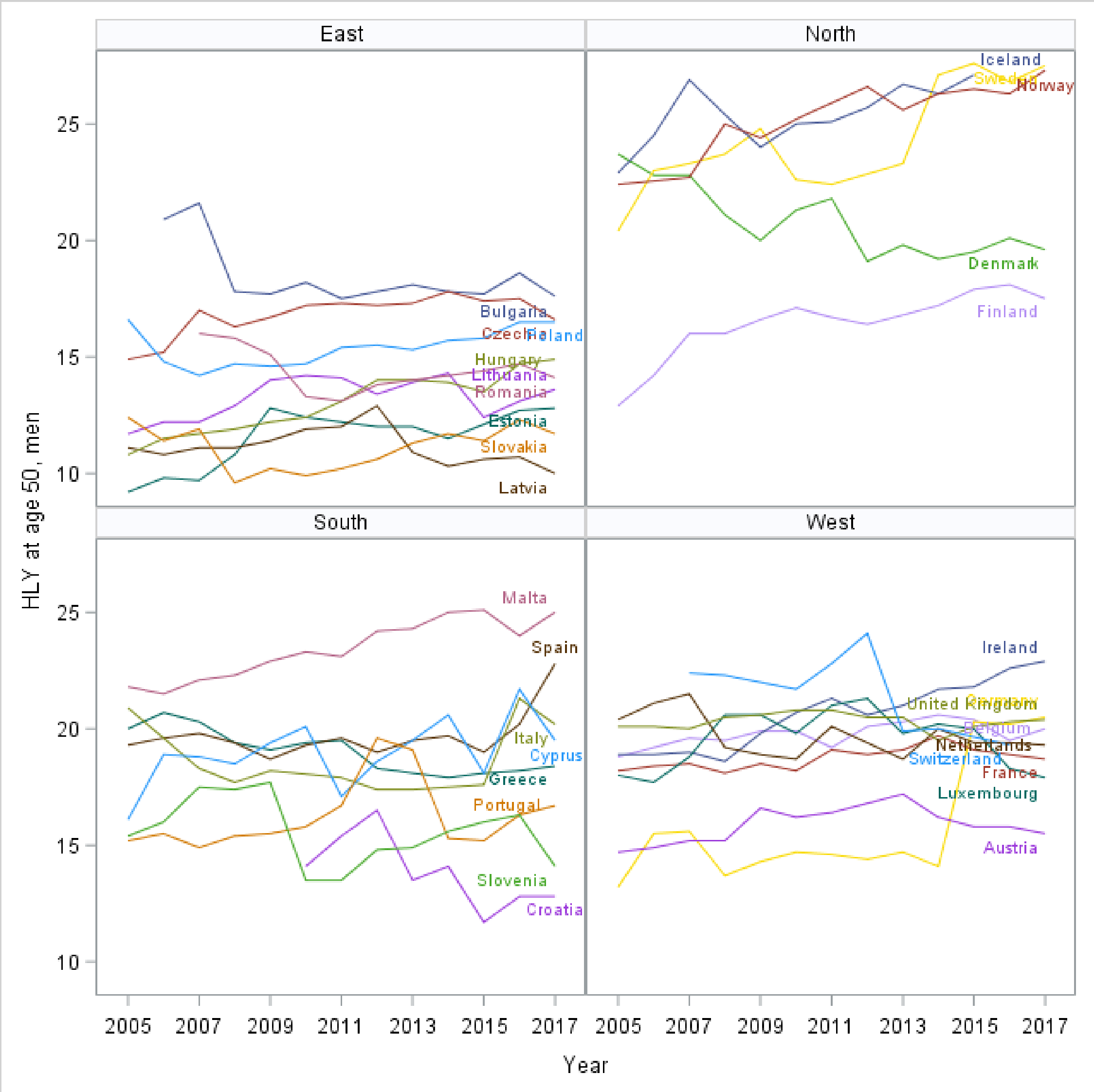
Source: Eurostat

Note: processed by SAS 9.4

The lack of clear trends for HLY is rather surprising, even more so when comparing HLY to life expectancy (LE). In the same countries and period, LE at birth and at ages 50 and 65 show for both sexes a clear upward trend; mean and median LE at age 65 increased by 1.7–2.0 years in the 12-year period, or by about 8–11 %. LE at birth increased by some 2.5–3.0 years or by 3–4 % in the period. In majority of countries, LE was increasing steadily in the period 2005–2017 (detailed data on life expectancy are available in Electronic Appendix).

Some clarification of this disparity is brought by analysis of HLY in the countries individually. Such analysis uncovers that trends in HLY in some countries are quite strongly influenced by sudden significant year-to-year changes in HLY values (Figures 2 and 3). These changes are unlikely to result from shifts in disability prevalence between two consecutive years, since health measures do not usually change with such speed. A more likely explanation is a shift in definition of the GALI question in SILC. Figures and comments correspond only to HLY at age 50 for both sexes. This specification is used so that it is possible to compare the results to those of Jagger et al. (2008) and Fouweather et al. (2015), who used HLY at age 50 as a central response variable in their studies. Charts for HLY at birth and at age 65 are available in Appendices 1 through 4, complete data for all specifications are available in Electronic Appendix.

Figure 3 – Healthy life years at age 50 in European countries, men, 2005–2017



Source: Eurostat

Note: processed by SAS 9.4

In some cases, such shift occurred only for one year – in which case, there is a steep increase followed by steep decrease, or vice versa. This happened in Latvia (HLY50 in 2012 for women was 1.7 years higher than in any other year; similar shift occurred with men, although less pronounced), Switzerland (HLY50 for men and women increased between 2011 and 2012 by 1.3 and 1.9 years, respectively), Portugal (where the shift lasted for two years, 2012 and 2013; difference to preceding and following years was up to 3.5 years for HLY50) or Iceland in 2007 (Figures 2 and 3).

In other cases, results show a permanent change in one year, which lasted thereafter and influenced values in all remaining years. An example of this case is Germany (where HLY50 for women, for instance, increased from 14.5 to 22.3 years between 2014 and 2015; values of HLY65 for women almost doubled between these two years) and partly also Switzerland, which therefore experienced both kinds of shifts (Figures 2 and 3).

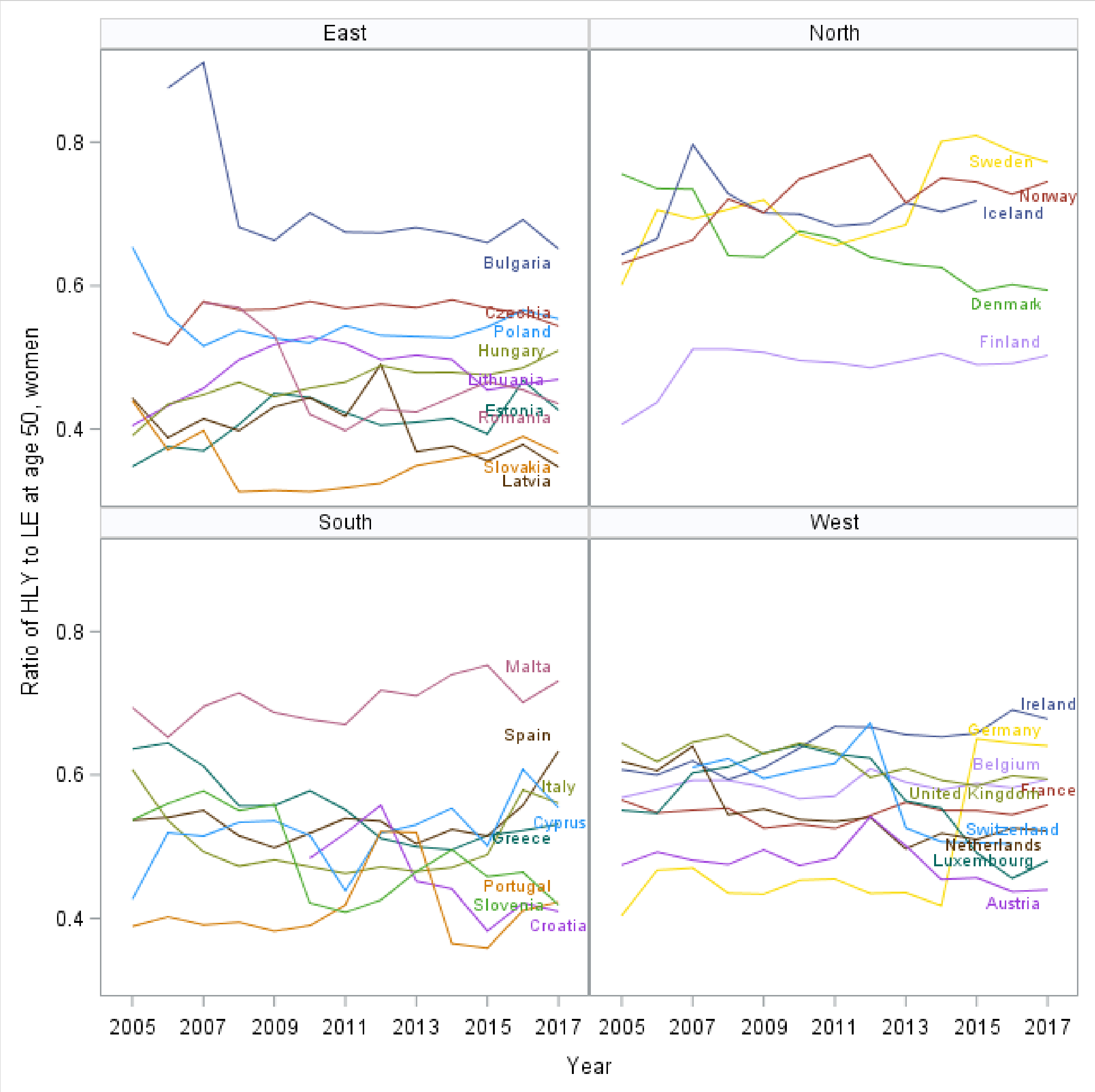
Reasons for these shifts can be in some cases connected to changes in GALI question wording: Germany experienced a significant change of the question in the 2015 survey (Eurostat, 2019d); in the years preceding 2015, question was not considered comparable with the standard English wording. Portugal changed the GALI question in 2012 and 2013; from 2014 on, the wording used in 2005–2011 was used again (Eurostat, 2019d). No corresponding changes in the question wording were found to explain the shifts in Latvia, Switzerland and Iceland.

Turning now to absolute values of HLY, it is clear that there are significant differences between European countries. Figures 2 and 3 display trends in HLY at age 50 for men and women, respectively, according to four European macro-regions. Countries were classified into the regions according to their location without any formal rules, only with an intention to make the charts as comprehensible as possible given the amount of data displayed.

It is clear that Scandinavian countries experience highest levels of HLY, and this region is the only one where a generally upward trend is observed (with the exception of Denmark, for which the values tend to decrease over time). In the rest of the countries, Malta is the only one that has HLY values comparable to those of Sweden, Norway and Iceland; and together with Hungary, Malta is also the only country that can stand comparison with Scandinavia in terms of rate of HLY increase. On the other side of the spectrum, one can find mostly Eastern or Central European countries – Estonia, Romania, Slovakia or Latvia. An exception within this region, which seems to be worth pointing out, is Bulgaria, whose HLY at age 50 exceeds values of a number of more developed countries, especially for women. Starting in 2008, the GALI question in Bulgaria is well comparable with the standard one, country‘s advance in HLY therefore is not caused by discrepances in definitions. Countries of Western and Southern Europe are somewhere in the middle, either slightly above or slightly below the European average.

Another interesting measure is the ratio of HLY to LE, which expresses the proportion of remaining life that is supposed to be lived without activity limitation. Ratios are shown in Figures 4 and 5 again for women and men at age 50, countries are regionally classified in the charts in the same way as with simple HLY. The shape of the curves generally corresponds to that of HLY, which is because LE increases relatively smoothly in virtually all analyzed countries throughout the years. Breaks in values of HLY are thus simply mirrored in values of the ratio of HLY to LE.

Figure 4 – Ratio of healthy life years to life expectancy at age 50, women, 2005–2017

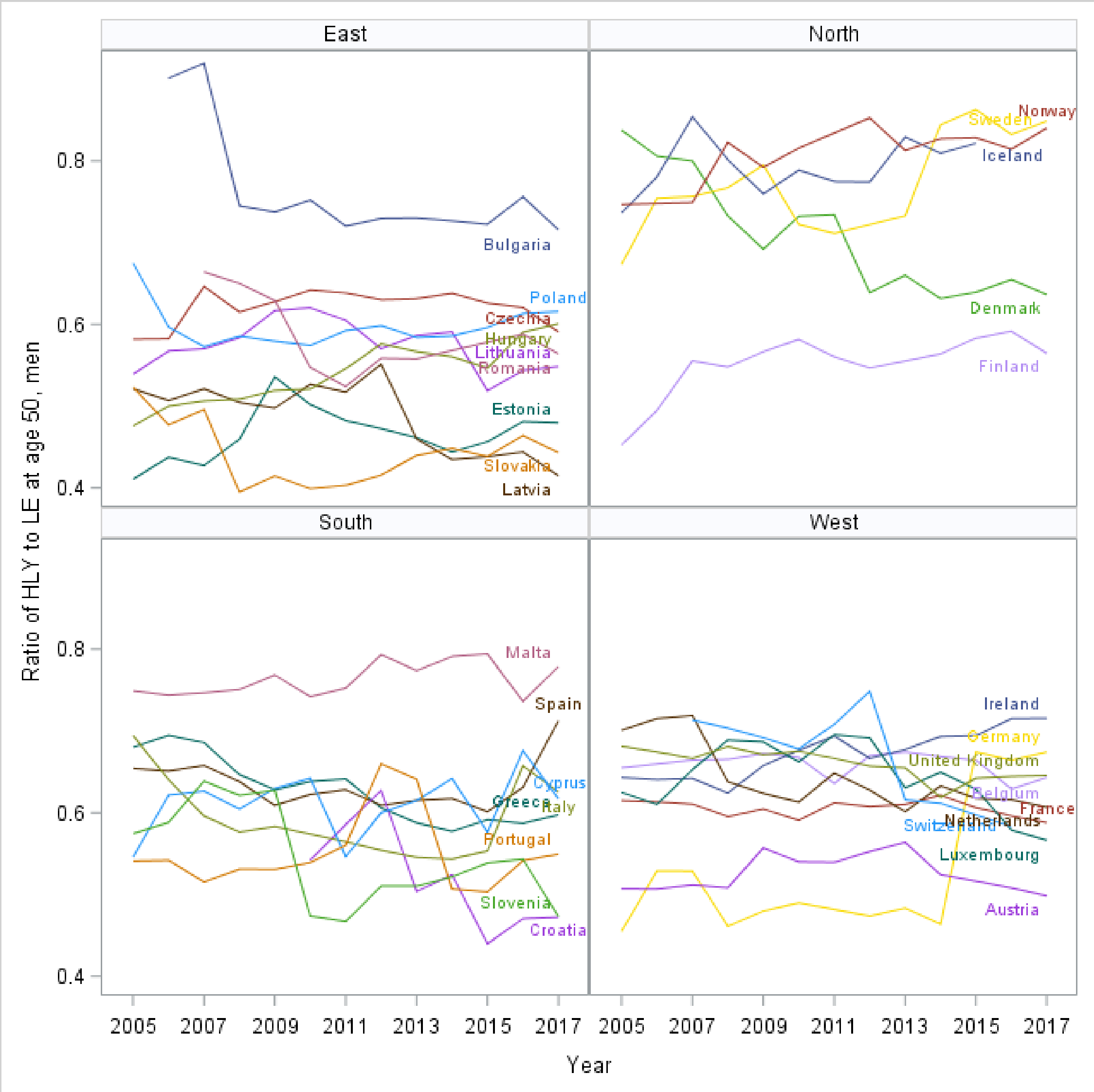


Source: Eurostat
Note: processed by SAS 9.4

This measure shows surprisingly high variation between countries. While people in some of the countries in the lead (including primarily Scandinavian countries together with Malta and Bulgaria) can at age 50 expect to live between 70 and 80 % of their remaining life free from activity limitation, inhabitants of Latvia, Slovakia, Portugal or even Austria are looking at only about 40 % of remaining life without activity limitation (Figures 4 and 5). This is certainly a wide range and in order to obtain an explanation, a detailed analysis would be necessary.

Moreover, values of the ratio of HLY to LE do not show signs of any convergence over time. This further reinforces the hypothesis mentioned earlier, that in terms of population health, European countries are not moving closer to one another. Jagger et al. (2013) warned in their analysis of European target, which was to increase HLY by 2 years by 2020, against moving attention from inter-country differences to global targets. They argued that setting global goals could contribute to an increase in inequality between countries (Jagger et al., 2013). Based on current HLY data, inequalities within Europe indeed did not diminish.

Figure 5 – Ratio of healthy life years to life expectancy at age 50, men, 2005–2017



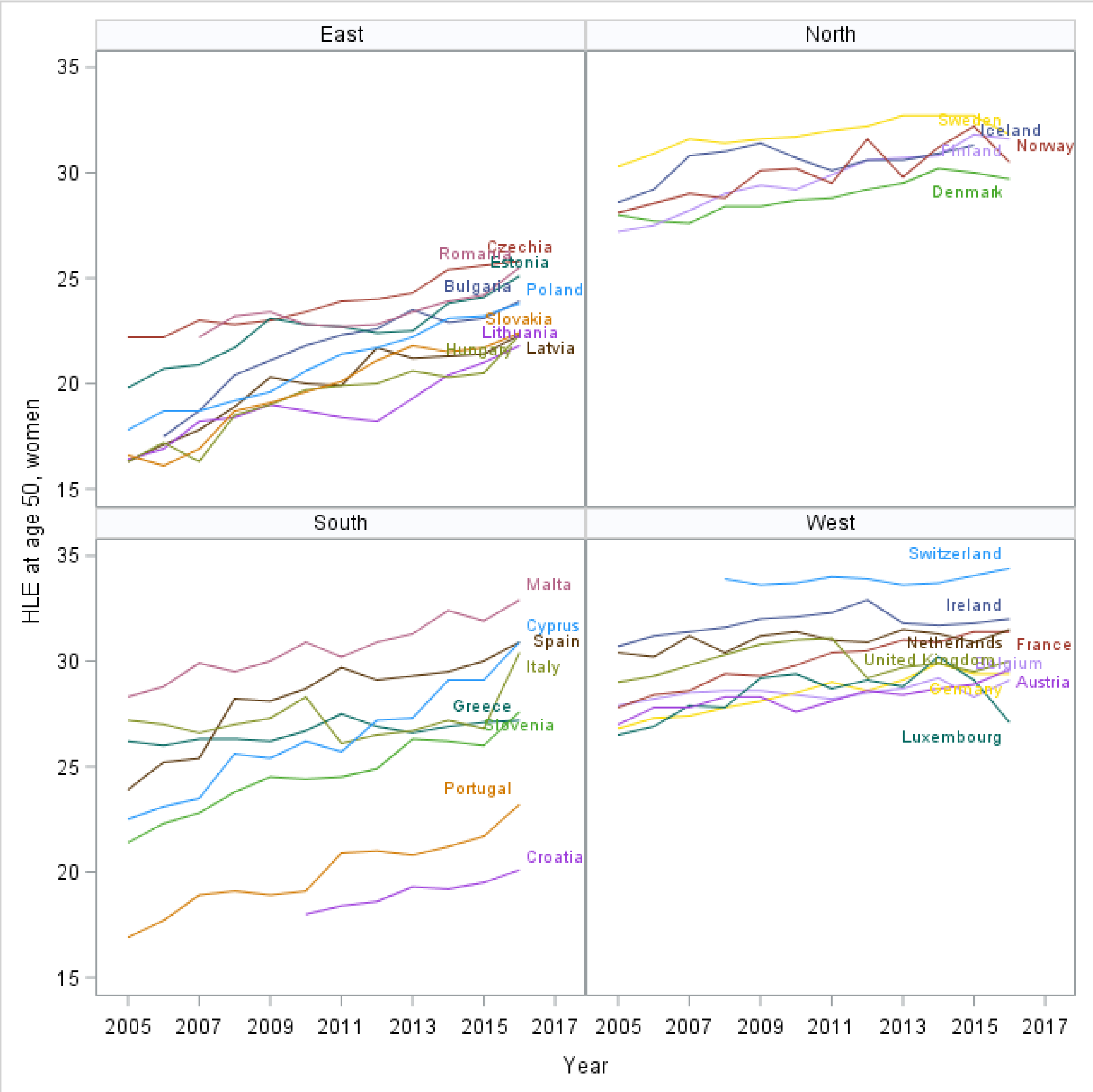
Source: Eurostat

Note: processed by SAS 9.4

5.1.2 Assessing healthy life expectancy based on self-perceived health

Healthy life expectancy based on self-perceived health (henceforth abbreviated as HLE) is based on a different measure of health, and adds therefore extra information to analysis of healthy life years. Instead of measuring how many years without limitation of everyday activities is a person on average expected to live, HLE expresses the number of years one is expected to spend in what is perceived as good health. HLE is not as dependent on the wording of the underlying question as HLY, and its values thus develop more steadily over time. This makes it easier to assess any trends that are present. As with HLY, comments in this section are related to HLE at age 50 for the same reasons as in section 5.1.1, but results do not fundamentally differ from HLE at birth or at age 65. Charts displayed here are restricted to women and serve as an illustration of the situation. Charts for HLE at age 50 for men, and at birth and at age 65 for men and women are displayed in Appendices 5 to 9.

Figure 6 – Healthy life expectancy based on self-perceived health at age 50 in European countries, women, 2005–2016

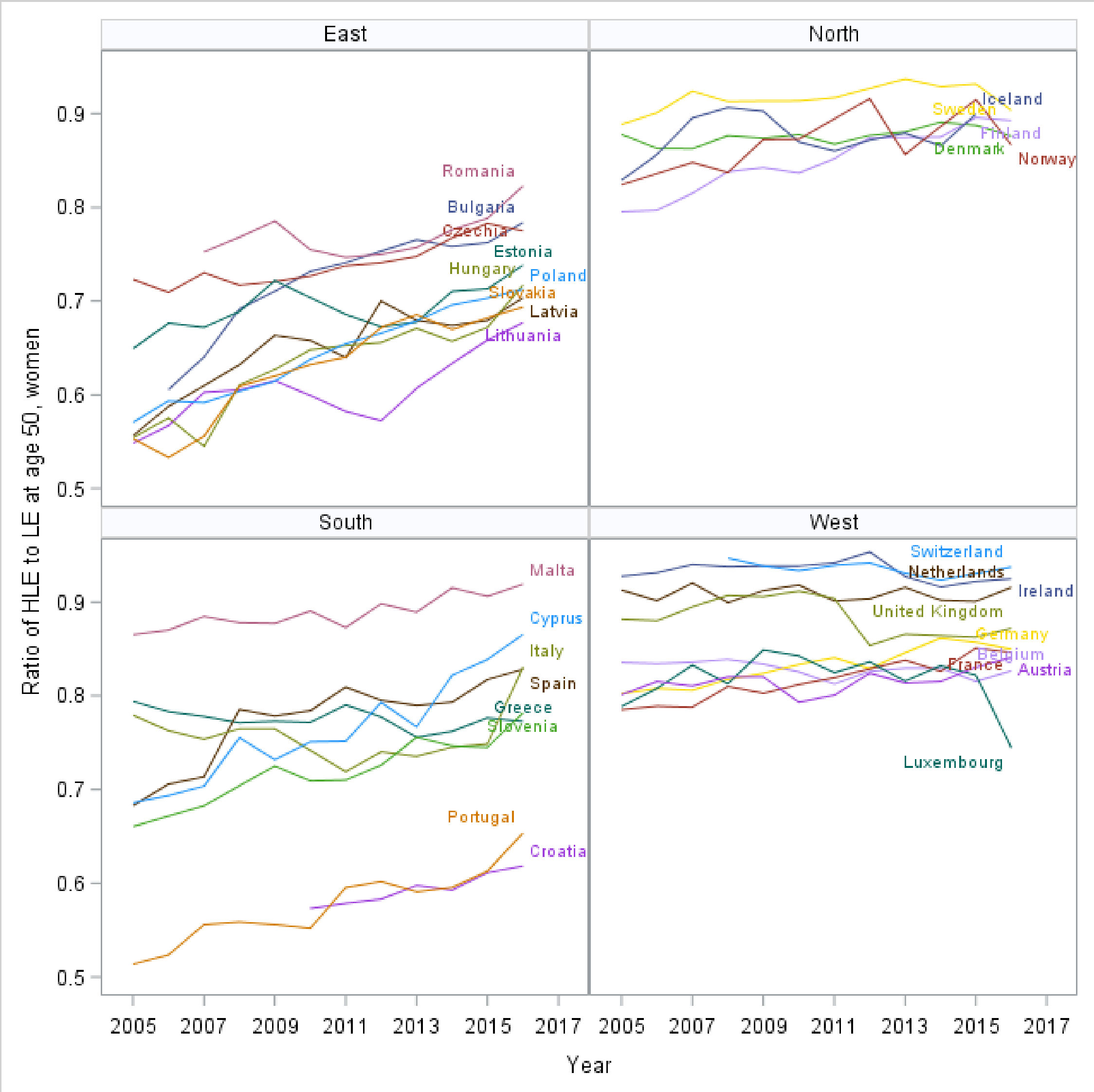


Source: Eurostat

Note: processed by SAS 9.4

First observation is that with HLE, Scandinavian countries are not as far ahead as with HLY. Values of HLE are similar across most of the wealthier countries of Northern and Western Europe – this statement can be in fact extended to countries that were members of the EU before 2004 (EU15), with the exception of Portugal, which is slightly lagging behind. Among these countries, Switzerland, Sweden and Iceland are the ones with highest HLE (Figure 6).

Figure 7 – Ratio of healthy life expectancy based on self-perceived health to life expectancy at age 50 in European countries, women, 2005–2016



Source: Eurostat
Note: processed by SAS 9.4

Second larger group of countries are countries in Central and Eastern Europe. These countries are also quite homogenous, with significant differences to the first group, but with relatively small differences between each other. These countries experience lower values of HLE, but are growing with a faster pace, which is a scenario that makes it likely that HLE values across the continent as a whole will be comparable in the foreseeable future (Figure 6). This is a finding that is in contrast with the results of HLY analysis – no convergence was observed for HLY.

Ratio of HLE to LE, as displayed in Figure 7, shows similarly high variation between countries as ratio of HLY to LE discussed in previous section. Patterns observed with absolute values of HLE are relevant for the ratio as well. In countries of Western and Northern Europe, ratio of HLE to LE is higher than in the rest of the countries, and is either growing slowly or stagnating. On the contrary, in countries of Central and Eastern Europe, values of the ratio are generally growing more rapidly. In 2005, these countries were lagging significantly behind the West, but the difference became much smaller in the course of the analyzed period. Portugal, Croatia and partly also Lithuania are outliers with lower values than most other countries.

Based on the comparison of HLY and HLE, and of their ratios to life expectancy, respectively, it seems in general that even years spent with activity limitation can be perceived as years that are spent in good health. HLE reaches in most cases higher values than HLY, which obviously holds also when expressed in relative terms as a proportion of life expectancy. This is in line with conclusion of Robine and Cambois (2013, p. 4), who write that „progress in life expectancy, shared by all EU countries, is associated, on average, with more years of […] activity limitations, but also with more years of perceived good health".

5.2 Trends in explanatory factors

The dataset contains a number of variables that are considered explanatory for purposes of further analysis in this thesis, and some general trends of these variables are briefly presented in this section. The goal of this section is not to provide the reader with a thorough analysis, so most of the time, focus is given to summary measures of the variables. Usually, for sake of brevity, one variable specification is chosen to represent one field – so, for instance, unemployment rate is discussed only in total, without the disaggregation to men and women. Such representative variables and their mean, minimum and maximum values in each year are shown in Table 3.

GDP presented here is standardized by purchasing power, which allows for price level differences between countries and is therefore suitable for cross-country comparisons. Summary statistics of GDP follow, unsurprisingly, overall economic cycles. The mean value was growing from 2005 until 2008, and decreased afterwards in connection with global economic recession. From 2009 on, values started to rise again, reaching values of 2008 no earlier than 2012. Overall, between 2005 and 2017, mean GDP increased from 23,700 to 31,400 Euros (Table 3).

The range between minimum and maximum values of GDP is wide, but is greatly influenced by high values of GDP in Luxembourg. Luxembourg is the only outlier, since its values reach often more than 1.5 times the maximum among other countries. There is no country with notably low GDP compared to the other ones.

Government expenditures on health increased overall between 2005 and 2017, but on average peaked in years 2009 and 2010. In these years, mean value was 6.4 % of GDP, while in 2005 it was 5.7 % and in 2017 6.1 %. The peak in the years of recession might be linked to decrease of GDP as a whole, and does not necessarily mean that absolute amount of money spent on health was higher in 2009 and 2010 than in other years. As with GDP, the difference between minimum and maximum values is considerable, 1.8–7.7 % of GDP in 2005 and 2.2–8.5 % of GDP in 2017 (Table 3).

Turning now to variables concerning income inequality, Gini coefficient of income distribution did not, on average, experience any significant change throughout the 12-year period. Its mean value was 29.3 in 2005 and 30.0 in 2017, with irregular fluctuations inbetween. Neither did minimum or maximum values experience any notable trends or shifts.

Table 3 – Selected summary statistics of representants of explanatory variables, 2005–2017

	year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
GDP (in purchasing power standard, in thousands)	Min	8.2	9.3	10.4	11.1	10.6	11.2	11.8	12.3	12.2	12.9	13.8	14.3	15.0
	Mean	23.7	25.3	27.0	27.2	25.2	26.1	27.0	27.6	27.7	28.7	30.4	30.5	31.4
	Max	57.9	64.5	69.1	68.4	62.5	65.4	69.3	69.2	70.0	74.4	78.0	78.4	78.0
Government expenditure on health (in % of GDP)	Min	1.8	1.7	1.7	1.8	1.9	1.9	1.9	2.1	2.2	2.1	2.2	2.2	2.2
	Mean	5.7	5.7	5.7	5.9	6.4	6.4	6.2	6.3	6.2	6.2	6.2	6.1	6.1
	Max	7.7	7.6	7.7	7.9	8.9	8.6	8.4	8.7	8.5	8.6	8.5	8.7	8.5
Gini coefficient of income distribution	Min	23.4	23.7	23.2	23.4	22.7	23.6	22.9	22.5	22.7	22.7	23.7	24.1	23.2
	Mean	29.3	29.6	29.4	29.6	29.5	29.4	29.3	29.2	29.5	29.9	30.0	29.8	30.0
	Max	38.1	38.9	38.3	37.5	37.5	37.0	35.1	35.7	35.4	35.6	37.9	37.7	40.2
At-risk-of-poverty or social exclusion rate (in %)	Min	13.3	12.5	13.0	11.8	11.6	13.7	13.7	12.7	13.0	11.2	12.5	12.2	12.2
	Mean	24.1	24.3	23.7	22.9	23.0	23.8	24.3	24.6	24.5	23.9	23.5	22.9	22.4
	Max	46.3	61.3	60.7	44.8	46.2	49.2	49.1	49.3	48.0	40.3	41.3	40.4	38.9
At-risk-of-poverty or social exclusion rate (65+, in %)	Min	6.4	6.4	7.2	5.4	6.2	5.3	4.5	5.0	4.2	6.4	6.1	6.4	8.6
	Mean	25.7	26.5	26.8	26.5	24.9	22.2	21.8	21.3	20.3	20.0	20.0	20.4	21.3
	Max	55.3	73.7	71.1	65.5	66.0	63.9	61.1	59.1	57.6	47.8	51.8	45.9	48.9
Population with low education (in % of total)	Min	10.1	9.7	9.5	9.1	8.6	8.1	7.1	6.7	6.6	6.7	6.5	5.4	5.2
	Mean	28.3	28.0	27.5	26.7	25.7	24.9	24.1	23.1	22.2	21.5	21.0	20.3	19.8
	Max	74.8	73.5	73.4	72.2	70.3	68.3	65.4	62.7	60.2	56.7	54.9	53.1	52.0
Population with tertiary education (in % of total)	Min	11.1	11.7	12.0	12.8	13.2	13.6	14.6	15.3	15.6	15.9	17.2	17.4	17.6
	Mean	23.4	23.9	24.6	25.5	26.5	27.3	28.2	29.2	30.2	31.6	32.3	33.1	33.8
	Max	34.6	35.1	36.4	36.6	37.3	38.7	39.3	40.8	42.6	45.9	44.5	45.1	46.5
Unemployment rate (in %)	Min	2.6	2.9	2.3	2.7	3.3	3.7	3.4	3.3	3.8	3.6	4.0	3.0	2.8
	Mean	8.1	7.2	6.3	6.2	8.7	9.9	9.8	10.4	10.7	10.1	9.3	8.4	7.3
	Max	17.9	13.9	11.2	11.3	17.9	19.9	21.4	24.8	27.5	26.5	24.9	23.6	21.5
Long-term unemployment rate (in %)	Min	0.8	0.8	0.4	0.3	0.4	0.7	0.7	0.6	0.7	0.6	0.5	1.2	0.3
	Mean	3.8	3.3	2.7	2.3	2.7	3.9	4.3	4.7	5.0	4.8	4.4	3.9	3.1
	Max	11.7	10.2	8.3	6.6	6.5	9.2	9.2	14.5	18.5	19.5	18.2	17.0	15.6

Note: Min and Max refer to minimum and maximum values of variables in respective year.

Source: Author's own calculations based on Eurostat

On the contrary, some decrease was observed with the at-risk-of-poverty or social exclusion rate, which is in Table 3 shown at all ages and at ages above 65 years. At all ages combined, mean value declined from 24.1 % to 22.4 %, but the decline was far from being steady – in 2012, for instance, the mean value peaked at 24.6 %. At ages above 65, mean at-risk-of-poverty or social exclusion rate grew between 2005 and 2007 from 25.7 % to 26.8 %, but decreased afterwards to values only slightly above 20 %. In spite of this overall decrease, there are countries which in the past years experienced substantial growth in terms of values of this rate – Latvia, Lithuania and Estonia. In these countries, together with Romania and Bulgaria, proportion of population older than 65 that is at risk of poverty or social exclusion is unfavourably high, reaching values over 40 %.

Educational structure is by definition something that changes slowly, but steadily. Summary statistics on proportion of population with low and tertiary education, respectively, are a proof of that. Mean proportion of population with low education shrinked between 2005 and 2017 from 28.3 % to 19.8 %. Correspondingly, proportion of tertiary-educated people in population increased by more than 10 percentage points, from 23.4 % to 33.8 % (Table 3). This trend was universally present in all studied countries.

The last two variables – unemployment rate and long-term unemployment rate – are characteristics of labour market and as such show a significant variation both over time and across countries. Until 2008, unemployment was declining, mean of standard and long-term unemployment rate reached its low in 2008 with values of 6.2 and 2.3 %, respectively. After that, unemployment returned back to growth and reversed again in 2013, after the impact of economic recession on job market faded away. In 2017, unemployment rate averaged at 7.3 % and long-term unemployment rate at 3.1 %. After 2008, the worst situation regarding unemployment was experienced by Southern European countries – Greece, Spain, Portugal and Italy. Highest values of unemployment rate in Greece and Spain even exceeded 25 % (Table 3).

Chapter 6

Methodology of regression analysis

The dataset that was described and briefly analyzed in previous two chapters is, by its nature, a *panel* dataset. This chapter aims to introduce the most important features of datasets that are structured in such a way, and thereafter discusses specific methods that are designed for panel data analysis so that all the features of such data are correctly accounted for. Finally, the most suitable methods for analysis are chosen and described before turning to reporting the results in Chapter 7.

6.1 General description of panel data and its features

The basic feature of the dataset for this thesis is a simple fact that there are two dimensions present – time and cross-sectional. However, not all datasets that have these two dimensions can be considered panel; in order to obtain a panel dataset, the same units are required to be followed across time (Wooldridge, 2018, p. 427). On the contrary, if a random sample of individuals is taken in different points in time, the resulting dataset is referred to as a pooled cross-section.

The nature of panel data brings a number of benefits, but – unsurprisingly – also some complications, which is why simple regression methods usually do not yield correct results when applied to panel data, and special models have thus been developed. The benefits are as follows: Firstly, panel data can control for individual heterogeneity – analyzed units can have some time-constant characteristics that differ across the units, but it is difficult or even impossible to measure them directly. These characteristics can be accounted for using panel data methods (Baltagi, 2005). Moreover, panel data have more variability (due to relatively high number of observations), less collinearity among variables and more degrees of freedom – which causes the estimates to have lower standard errors (Baltagi, 2005).

Specific complications arising with panel data analysis include dealing with unobserved characteristics of cross-sectional units, which are the same across time. In fact, in order to achieve the first benefit (controlling for individual heterogeneity), the bias which would arise when using standard cross-sectional methods for panel data has to be removed by using special methods (Baltagi, 2005). Apart from that, general issue with panel data can be the difficulty to collect such a dataset, since following the same individuals across time is not always simple (Wooldridge, 2018). Fortunately, given that units in this thesis are countries which are, from a statistical point of view, well covered by Eurostat, this issue does not need to be handled here.

On the other hand, an issue that must be paid attention to is missing data. On the basis of occurrence of missing data, a panel is classified either *balanced* or *unbalanced*. A balanced panel is one that has observations for every variable at all points in time. The total number of observations is therefore nT, where n stands for number of cross-sections and T stands for number of time periods (Park, 2011). An unbalanced panel, on the contrary, contains missing values for at least one variable. The total nuber of observations in an unbalanced panel is therefore less than nT (Park, 2011).

It is clear from previous chapters that the dataset for this thesis is unbalanced – it does not contain a complete set of observations. However, standard panel data methods can be applied to an unbalanced panel as well, with some minor limitations (Park, 2011). The occurrence of missing data would constitute a problem only in case the missing values were systematic, which appears not to be the case (based on the discussion in Chapter 4). With missing data occurring randomly, analysis can be performed in a usual way without any implication for the resulting estimates (Wooldridge, 2018).

6.2 Regression methods for panel data

As was already pointed out, the purpose of panel data methods is to account for the unobserved fixed (or individual) effects that are often by definition present in panels. Suppose the general regression model for one explanatory variable x is constructed in the following way, where y is the response variable, β_0 and β_1 are the regression coefficients and i and t are subscripts for cross-sectional units and time, respectively (based on Wooldridge, 2018, p. 459):

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}$$

In this equation, the sum $a_i + u_{it}$ is the disturbance term, expressing all the variation in y not explained by x. The first part of the disturbance term, a_i, is the fixed (also called individual or unobserved) effect. It is the unmeasured characteristics of a cross-sectional unit that is, for one specific unit, constant over time. The other term, u_{it}, is a time-varying error, in literature referred to as idiosyncratic error (Wooldridge, 2018). Idiosyncratic error captures all factors that affect y, are not included in x and vary over time.

There are four basic approaches to running linear regression on panel data: (1) pooling ordinary least squares (OLS) across time, (2) running OLS with first-differencing, (3) using fixed effects (FE) estimation and (4) using random effects (RE) estimation. In the rest of this section, these four approaches are described and discussed using mainly econometric textbooks by Baltagi (2005) and Wooldridge (2018) and a tutorial paper by Park (2011).

The first and simplest method is pooling OLS across time, and thus treating the dataset as if it was pooled cross-section. Pooled OLS is suitable in cases where it is reasonable to suppose that the fixed effect is non-existent. In such a case, disturbance term does not suffer from serial correlation and under standard OLS assumptions, resulting estimates are unbiased and consistent (Park, 2011; Wooldridge, 2018). However, panel data usually do not satisfy the core condition for pooled OLS.

6.2.1 First differencing

Except for pooling OLS, all the other methods are based on the presence of fixed effect, but differ in the specific approach of dealing with it. The idea behind first differencing is that if a regression equation for time *t-1* is subtracted from the equation for time *t*, then the resulting differenced equation does not contain the fixed effect, since it is the same in both time periods (Wooldridge, 2018). Any other time-constant variable is also differenced-away together with the fixed effect. The resulting equation (using, for purpose of illustration, only a single explanatory variable and differencing between two time periods), is

$$\Delta y_i = \delta_0 + \delta_1 \Delta x_i + \Delta u_i$$

Here, δ coefficients are used to underline the fact that they are based on estimation of a differenced equation. This equation can be, under standard assumptions, estimated by OLS (Woodridge, 2018). Method can be arbitrarily extended to more than two time periods and multiple explanatory variables.

Usage of first differencing is practically limited only to balanced data panels, since the method requires observations in both adjacent time periods to be non-missing in order to compute their difference. If the panel is unbalanced, the number of differences that can be possibly computed decreases, and although first differencing can be used even in these cases, other methods (mainly fixed effects) are more appropriate (Wooldridge, 2018).

6.2.2 Fixed effects estimation

Fixed effects estimation and random effects estimation are used more often in applied work than the preceding two approaches. Pooled OLS and first differencing are simple to understand and apply, but their simplicity is offset with a number of assumptions that need to be met. Both FE and RE are designed to be used with the presence of individual effect (which makes them usually preferable over pooled OLS) and, unlike first differencing, can be used even with unbalanced panels (Park, 2011; Wooldridge, 2018).

The logic behind fixed effects approach is not too dissimilar to that of first differencing, and with two time periods, the two methods are completely identical. The difference is that fixed effects estimation subtracts time-average from each variable in the regression (Wooldridge, 2018). If the basic model is constructed in the same way as in the beginning of Section 6.2, i.e.

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it},$$

the first step in order to obtain the fixed effects transformation is to average the equation over time for each *i* (Wooldridge, 2018, p. 484):

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + a_i + \bar{u}_i .$$

In this equation, $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$, where T denotes total number of time periods. The remaining averages are constructed analogously. Since intercept and individual fixed effect are constant over time, it makes no sense to construct an average of them. Finally, the fixed effects transformation requires subtracting the averages from the original equation, giving a time-demeaned equation (Wooldridge, 2018, p. 485)

$$y_{it} - \bar{y}_i = \beta_1(x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i .$$

This equation can be estimated by OLS, and the resulting estimators (fixed effects estimators) are unbiased under the assumption that idiosyncratic error u_{it} is uncorrelated with explanatory variables in all time periods. Other OLS assumptions require the error to be homoskedastic and serially uncorrelated, and if these assumptions are met, the fixed effects estimator is best linear unbiased estimator (BLUE; Baltagi, 2005).

FE estimation allows the fixed effect to be correlated with any or all of the explanatory variables (Park, 2011), which is often the case in practice. Time-demeaning causes the effect to disappear from the transformed equation, so the correlation does not cause troubles for statistical inference. As with first differencing, the fixed effects approach cannot evaluate the effect of any variable that is fixed over time, since such a variable would be differenced away in the transformation (Woodridge, 2018).

6.2.3 Random effects estimation

In cases where it is reasonable to assume that individual effect is uncorrelated with all of the explanatory variables, random effects estimation can be used (Baltagi, 2005). In RE approach, the goal is not to eliminate the individual effect a_i, because if it is uncorrelated with explanatory variables, its omission will result in inefficient estimators. However, if a_i is included in the error term and at the same time is constant over time for each cross-section, the error term is serially correlated (Wooldridge, 2018).

Random effects transformation eliminates this serial correlation by so-called quasi-demeaning of the original equation. The transformed equation, again with one explanatory variable for purpose of simplicity, is (Wooldridge, 2018, p. 493)

$$y_{it} - \theta\bar{y}_i = \beta_0(1-\theta) + \beta_1(x_{it} - \theta\bar{x}_i) + (v_{it} - \theta\bar{v}_i) .$$

In this equation, v_{it} is the composite error term $a_i + u_{it}$ and θ is a parameter between 0 and 1, constructed on the basis of variances of the two components in composite error term and the number of time periods T (Wooldridge, 2018, p. 493):

$$\theta = 1 - \sqrt{\sigma_u^2 / (\sigma_u^2 + T\sigma_a^2)} ,$$

where σ_u^2 is the variance of u_{it} and σ_a^2 is the variance of a_i. Compared to fixed effects estimation, random effects does not subtract whole time-average, but only a fraction of it, where the fraction is defined by θ. The exact value of θ is never known, but can be estimated. This subtraction eliminates the serial correlation of the composite error term, and the resulting quasi-demeaned equation can be estimated by OLS (Wooldridge, 2018).

Using RE estimation is helpful when the purpose of analysis is to evaluate effect of a time-constant variable (Wooldridge, 2018), but in many applications, the key assumption for RE (no correlation between unobserved effect and explanatory variables) is rather strong. However, in case it holds, RE estimator is more efficient than an estimator obtained by any of the other three methods discussed (Wooldridge, 2018).

6.3 Choice of methods and overview of steps of analysis

6.3.1 Choice of general approach

Specific definition of the models that will be estimated and reported follows in the subsequent section of this chapter, but in order to be able to decide on methods that will be employed, a general framework for the model must be set. The model construction simply follows the goals and hypotheses that were set in introductory chapters.

Since the central topic in this text are health expectancies, it is without surprise that response variable in all the models will always be a measure of health expectancy. Explanatory variables will be GDP, expenditures on health care, a measure of education, standard or long-term unemployment rate and a measure of inequality in income distribution.

In panel data analysis, as was introduced in the previous section, it is the unobserved individual effect that is of crucial importance when deciding which approach to choose. Since the cross-sections in this study are countries, it is reasonable to assume the existence of an unobserved individual effect. Health expectancy is supposedly affected by a set of cultural and social norms and values which are specific for individual countries, but which cannot be directly measured – and this thesis does not search for proxy variables for these phenomena (link between characteristics of social values and institutional settings of a country and measures of population health was found e.g. by Minagawa (2013)). The unmeasured norms and values are therefore represented by the unobserved individual effect in the model. It is likely that characteristics represented by the unobserved effect are correlated at least with some of the explanatory variables – GDP, unemployment or educational structure. This assumption thus rules out the random effects approach to model estimation.

One more issue to discuss before an approach is chosen is the fact that the dataset is unbalanced – there is a number of variables that have missing values. Since their distribution seems to be random and not correlated with idiosyncratic errors, this fact does not hamper the analysis. However, it helps when deciding on the most suitable approach.

Based on these issues regarding the dataset, fixed effects estimation is supposed to be the best option for evaluating the effects of selected macro-level variables on health expectancy. Fixed effects estimation is thus considered a basic attitude towards the analysis, but its validity will nevertheless be tested, as is described in the following paragraphs.

6.3.2 Specification of models

Before turning to the process of model specification, it is useful to establish a framework according to which the variables are named, so that any further referencing and reporting is easier. Table 4 summarizes the technical names for variables and reports their description. The technical names are hereafter used while describing the models and reporting their results. The letter *f* at the end of variable name means the variable is specified for women; letter *m* stands for men.

Table 4 – Overview of technical names of variables

Variable name	Variable description
hly0, hly50, hly65	Healthy life years (HLY) at age 0, 50, 65
hle0, hle50, hle65	Healthy life expectancy (HLE) based on self-perceived health at age 0, 50, 65
GDP, GDP_PPS	GDP in purchasing power standard
GDP_euro	GDP in Euros
hlthexp	Government expenditure on health
Inc_Gini	Gini coefficient of income inequality
povrisk	At-risk-of-poverty or social exclusion rate (AROPE) at all ages
povrisk65	At-risk-of-poverty or social exclusion rate (AROPE) at ages 65+
educ_high	Proportion of population with tertiary education
educ_low	Proportion of population with low education
unem	Unemployment rate
unem_long	Long-term unemployment rate

Source: author's overview

As a first step, correlations among explanatory variables are analyzed in order to avoid possible multicollinearity. Wooldridge (2018, p. 90) defines multicollinearity as a situation where high (but not perfect) correlation exists between two or more explanatory variables. There is no exact threshold above which correlation is supposed to be „high“; instead, evaluation is left to the analyst (Wooldridge, 2018).

One of the assumptions for multiple linear regression to yield valid estimates is that there exists no perfect collinearity (no exact linear relationships) among explanatory variables (Wooldridge, 2018). Cases when correlation coefficients between variables are high, but not equal to one, therefore do not violate this assumption. However, in the presence of multicollinearity, inference can become inaccurate due to high variance of estimators. As a consequence, incorporating different variables with a high degree of mutual correlation in the model is implausible.

Pairwise correlations among possible explanatory variables are displayed in Table 5. Table shows values of Pearson coefficients of correlation for variables related to women (where sex distinction is applicable), but results for men are effectively the same (correlation matrix for men is reported in Appendix 10). As is expected, there is a strong correlation between GDP expressed in Euros and GDP expressed in purchasing power standard, and also between standard unemployent rate and long-term unemployment rate.

Table 5 – Pairwise Pearson coefficients of correlation between variables, women (where applicable)

	GDP_PPS	GDP_euro	Hlthexp	Inc_Gini	povrisk_f	povrisk65_f	educ_high_f	educ_low_f	unem_f	unem_long_f
GDP_PPS	1									
GDP_euro	0.97	1								
Hlthexp	0.28	0.36	1							
Inc_Gini	-0.39	-0.41	-0.49	1						
povrisk_f	-0.58	-0.59	-0.61	0.75	1					
povrisk65_f	-0.61	-0.59	-0.59	0.61	0.79	1				
educ_high_f	0.37	0.43	0.07	-0.06	-0.23	-0.10	1			
educ_low_f	0.02	0.01	0.03	0.15	0.01	-0.08	-0.44	1		
unem_f	-0.34	-0.36	-0.13	0.35	0.35	0.05	-0.10	0.15	1	
unem_long_f	-0.37	-0.39	-0.13	0.32	0.38	0.07	-0.22	0.11	0.94	1

Source: own calculation based on data from Eurostat

Moreover, the triplet of variables Gini coefficient of income distribution, at-risk-of-poverty or social exclusion rate for all ages combined and for ages above 65 is correlated with values of correlation coefficients around 0.7. Since all these three measure similar phenomena, it is reasonable to ony include one of them in the models. Based on the correlation matrix, it is – quite surprisingly – possible to include both proportion of population with low education and with tertiary education, the correlation between these two is not substantial.

Apart from that, GDP variables are somewhat correlated to a number of other variables. But since none of these pairs (GDP and education, GDP and poverty risk) cannot be considered to measure similar phenomena, dropping any of these variables would not be without complications.

Based on considerations that were made up to this point in the text, first model (hereafter referred to as Model 1F) is formulated in the following way:

Model 1F

$$\text{hle0\_f} = \beta_1\text{GDP} + \beta_2\text{hlthexp} + \beta_3\text{povrisk65\_f} + \beta_4\text{educ\_high\_f} + \\ \beta_5\text{educ\_low\_f} + \beta_6\text{unem\_long\_f} + a_i + u_{it}$$

A convention that is held throughout the rest of this text is that the letter following the model number is *F* for women and *M* for men – Models 1F and 1M are thus identical, except for gender specification. For sake of better comprehensibility of the results, *GDP* values entering the regression are expressed in thousands. The same holds in models 3 and 4 which are presented further on.

There are two comments towards the model formulation. Firstly, the response variable is healthy life expectancy based on self-perceived health at age 0. This variable was chosen arbitrarily out of the three ages for which values are collected, and there is no model for HLE at ages 50 or 65. That is because HLE at ages 0, 50 and 65 are almost perfectly correlated, and all conclusions derived for one of them are for this reason valid also for the other two. Secondly, HLE is chosen as the principal response variable as opposed to HLY for reasons related to data quality and comparability, that were in detail discussed in chapters 4 and 5.

In literature examining various socio-economic factors and their impact on health expectancies, little support is found for using other than simple linear functional forms for explanatory variables in modelling process (e.g. Gutierrez-Fisac et al., 2000; Jagger et al., 2008; Fouweather et al., 2015). One exception is Kabir (2008), who includes explanatory variables in logarithmic form, having life expectancy as the response variable.

In model 1, all variables are included without transformation. However, after visual assessment of the scatter plot of HLE and GDP in purchasing power standard (Appendix 11 and 12), some support is found for a possible log-linear relationship between these variables. In order to examine this assumption formally, model 2 is formulated in the same way as model 1, but natural logarithm of GDP (log(GDP)) is included instead of the standard form. Logarithmic transformation allows for non-linear relationship between variables, and if this is the case in reality (which can never be known), model estimates the parameters more precisely. Specification of model 2 with variables for women is as follows:

Model 2F

$$\text{hle0\_f} = \beta_1 \log(\text{GDP}) + \beta_2 \text{hlthexp} + \beta_3 \text{povrisk65\_f} + \beta_4 \text{educ\_high\_f} + \\ \beta_5 \text{educ\_low\_f} + \beta_6 \text{unem\_long\_f} + a_i + u_{it}$$

The idea for further models is based on Jagger et al. (2008) and Fouweather et al. (2015) and intends to run the analysis separately for several groups of countries. Both these authors studied two groups of countries within EU, first one consisting of EU members that joined EU before 2004 (referred to as EU-15), and second one consisting of those that entered EU in 2004 (referred to as EU-10). Romania, Bulgaria and Croatia, joining EU after 2005, were not fully covered by these two studies.

Models 3 to 6 follow, for purpose of comparison, this distinction to some extent. Countries are split in two groups, where the basic criterion is whether they joined EU before 2004 or starting by 2004. Those countries that are not EU members, are all assigned to group together with countries that were EU members before 2004. First group, also called further on „old member countries“, thus consists from EU-15, plus Switzerland, Norway and Iceland. Second group consists of those that joined EU in 2004, 2007 (Romania and Bulgaria) and 2013 (Croatia).

Models 3 and 4 are, as for variable specification, identical to model 1, and differ with respect to countries included. Model 3 analyzes old EU members, while model 4 analyzes only new EU member states. The same concept is used for models 5 and 6, which are based on the specification of model 2, thus using logarithmic transformation of *GDP*. As for countries included, model 5 uses old EU members, model 6 is based on new member states.

6.3.3 Steps of analysis

The analysis is performed in SAS software, whose features enable the user to perform many kinds of statistical analyses. In this text, the most important procedure is PANEL procedure, that is able to compute estimates based on fixed effects approach, random effects approach, first differencing and many others (SAS Institute, 2014). PANEL procedure also provides several tests for evaluation of the model and its properties (SAS Institute, 2014), some of which are described further on in this section.

In order for the estimators of parameters specified in the models to have desired properties, certain assumptions need to be met. One of them, no perfect collinearity, has already been discussed. A key assumption for a fixed effects estimator to be unbiased (definition of unbiasedness is provided e.g. in Wooldridge, 2018) is strict exogeneity. Because the presence of unobserved, time-invariant effect is assumed in the models, the idiosyncratic error is assumed to be uncorrelated with explanatory variables, which means the strict exogeneity is supposed to hold.

Under additional conditions of heteroskedasticity and absence of serial correlation, the estimator is BLUE – best linear unbiased estimator. Since in SAS, testing for heteroskedasticity and serial correlation in unbalanced panels is far from being straightforward, the approach chosen in this respect is to employ Newey-West estimator and to report standard errors that are corrected for possible heteroskedasticity and serial correlation (according to Newey and West, 1987). The conclusions concerning statistical significance of estimates are not substantially influenced by this adjustment.

When running FE regression, PANEL procedure in SAS routinely computes F-test as a specification test in order to determine whether there are any fixed effects in the model. Rejecting the null hypothesis of no fixed effects is considered to be a basic condition for using FE approach (Park, 2011). Results of F-test are not usually reported in the remainder of the text and the null is implicitly considered to be rejected, unless explicitly stated otherwise. In such a case, it is problematic to estimate the model by FE approach and another approach would have to be chosen – however, this is not the case for any model estimated and reported in the following chapter.

In applied panel data analysis, Hausman test is often used to decide whether FE or RE approach is more suitable. If the null hypothesis of Hausman test is rejected, FE is favoured over RE. However, Wooldridge (2018) points out that failure to reject the null does not necessarily mean that RE should be used, but that there can be other issues that can play their role. Hausman test thus is not reported, and the reasoning from section 6.3.1, which supports fixed effects, is used instead.

In view of all the issues discussed in this subsection or earlier, the only outcomes that are reported and commented on in chapter 7 are estimates obtained by fixed effects regression of the models, together with standard errors of the estimates and resulting p-values, where the standard errors are robust to heteroskedasticity and serial correlation thanks to employing Newey-West method. Since null hypothesis of F-test for fixed effects is in all cases rejected on 5% level, there is no need to report any specification tests, as was anyway already justified.

Chapter 7

Results of analysis

This chapter reports the results of models that were presented and specified in chapter 6. Attention is paid to the estimates of parameters, their practical and statistical significance, and to goodness-of-fit of the models. After having described the model outcomes, second section of this chapter turns to discussion of the results. In that part, results are generalized and conclusions are confronted with theoretical assumptions and with literature covering the topic.

7.1 Outcomes of regression models

7.1.1 Models for all countries combined

First and basic model regresses HLE at age 0 on a set of six explanatory variables – GDP in purchasing power standard, government expenditure on health, at-risk-of-poverty or social exclusion (AROPE) rate for people aged 65 or more, proportion of population with tertiary education, proportion of population with low education and long-term unemployment rate. Except for GDP and government expenditure on health, variables are differentiated by sex. Model 1F includes variables for women, model 1M includes variables for men. GDP and government expenditure on health are used in their general form in both models. In models 1F and 1M, explanatory variables are used without any transformation and the parameters resulting from estimation of these models (together with standard errors (SE) of the estimates and p-values) are reported in Table 6.

First remark is that both models yield very high goodness-of-fit, represented by R-squared. Values 0.956 and 0.971, respectively, would in a standard OLS regression indicate that the model explains variation in the response variable almost perfectly. In a fixed effects estimation, however, this high values of R-squared are not unusual and should be therefore interpreted with caution (Wooldridge, 2018). Rather than evaluating the absolute value of R-squared, it is more informative to use it for mutual comparison of different models estimated by the same method.

Table 6 – Parameter estimates and other results of models 1M and 1F

	Model 1F			**Model 1M**		
	Estimate	SE	p-value	Estimate	SE	p-value
GDP	-0.024	0.047	0.6116	0.187	0.042	<.0001
hlthexp	-0.019	0.175	0.9151	0.566	0.150	0.0002
povrisk65	-0.081	0.020	<.0001	-0.089	0.016	<.0001
educ_high	0.228	0.047	<.0001	0.093	0.048	0.0533
educ_low	-0.025	0.028	0.3641	-0.078	0.030	0.0101
unem_long	-0.113	0.049	0.0219	0.060	0.036	0.0926
R-squared	0.956			0.971		

Source: output of regression analysis performed in PANEL procedure in SAS 9.4

The parameter estimates and their significance differs somewhat between models 1F and 1M. When evaluating the results, 0.05 level of significance is taken as a threshold in deciding whether or not the estimate is significantly different from zero. The 5% level is used most often in econometric analyses (Wooldridge, 2018) and is also used by other studies on similar topics (e.g. Gutierrez-Fisac et al, 2000; Jagger et al., 2008; Fouweather et al., 2015).

In model 1F, three parameter estimates are significant on 5% level – *povrisk65*, *educ_high* and *unem_long*. The signs of the parameters are as expected – negative for poverty risk and long-term unemployment, and positive for proportion of tertiary educated people. Increase of one percentage point in *educ_high* is associated with an increase of about 0.22 years in HLE at birth. The same increase in *povrisk65* and *unem_long* is associated with a decrease in HLE by about 0.08 and 0.11 years, respectively. Parameters for *GDP*, *hlthexp* and *educ_low* are not significantly different from zero in this model specification (Table 6).

Model 1M yields slightly higher R-squared than model 1F (0.971 compared to 0.956), and indicates significant association of HLE at birth with GDP and health expenditures in a positive direction, and with AROPE rate and proportion of population with low education in a negative direction. *Educ_high* has a positive coefficient on the margin of significance (p-value = 0.053).

The only variable, for which the parameter was in both 1M and 1F estimated consistently, is *povrisk65*. In other cases, differences in inference are substantial – most visibly in the case of *GDP* and *hlthexp*, which are both very significant (and in expected directions) in model for men, but far from any conventional significance level for women.

Second model differs from the first one only in functional form of GDP – the natural logarithm of GDP is used. Apart from that, specification remains the same. As can be seen in Table 7, results of models 2M and 2F differ to some extent from those of models 1M and 1F. Looking at R-squared, which is higher in 2M and 2F than in the previous models (although magnitude of the difference is negligible), one can conclude that changing functional form of GDP from linear to logarithmic was helpful for goodness-of-fit of the model.

Unlike in models 1M and 1F, *log(GDP)* is now significant for both men and women, with parameter estimates being equal to approximately 4.7 for women and 7.3 for men. These values mean that, holding everything else constant, 1% increase in GDP will result in HLE higher by about 0.047 and 0.073 years.[3] Long-term unemployment is insignificant in model 2F (p-value = 0.621) as opposed to the previous model. Otherwise, conclusions derived from model 2F are practically the same as in model 1F. The estimates show significant positive association of HLE with *educ_high* and significant negative association with *povrisk65*, the estimated parameter is in both cases closer to zero than in model 1F. With the inclusion of GDP in logarithmic form, both *educ_high* and *povrisk65* are therefore expected to have weaker relationship with HLE.

Table 7 – Parameter estimates and other results of models 2M and 2F

	Model 2F			**Model 2M**		
	Estimate	SE	p-value	Estimate	SE	p-value
log(GDP)	4.732	1.295	0.0003	7.342	0.912	<.0001
hlthexp	0.248	0.184	0.1792	0.549	0.139	<.0001
povrisk65	-0.073	0.019	0.0001	-0.054	0.018	0.0022
educ_high	0.120	0.043	0.0058	0.067	0.037	0.0739
educ_low	-0.036	0.026	0.1626	-0.069	0.029	0.0190
unem_long	0.026	0.052	0.6210	0.140	0.036	0.0001
R-squared	0.959			0.976		

Source: output of regression analysis performed in PANEL procedure in SAS 9.4

Model 2M indicates significant association of HLE with all explanatory variables except for *educ_high*, and even that is not very far from 0.05 significance level (*educ_high* has p-value equal to 0.074). Setting aside *log(GDP),* which was already commented on, three out of four remaining significant estimates are in the expected direction – *hlthexp*, *povrisk65* and *educ_low.* The estimate on *hlthexp* is quite high – one percentage point increase in government expenditure on health is modelled to be associated with an incrcease in HLY of 0.549 years. Fourth variable, *unem_long*, is predicted by the model to be positively associated with HLE, which is certainly not an expected result. According to model 2M, one percentage point increase in long-term unemployment rate would result in 0.14 years of increase in HLE.

7.1.2 Models for subsets of countries

For purpose of models 3 to 6, countries were split in two groups, that are for simplicity called „old EU countries“ and „new EU countries“. Those countries that are not EU members were assigned to the two groups on the basis of their regional characteristics, as was in detail described in previous chapter.

Models 3 and 4 are functionally specified in the same way as model 1. Model 3 takes into account only old EU countries, whereas model 4 analyzes new EU countries. Parameters estimated by both models are displayed in Tables 8 and 9, respectively. The two sets of estimates differ in a number of substantial respects, suggesting that splitting the set of coutries

[3] These values are calculated as the value of the estimate divided by 100. This is an approximation which works well for very small increases in explanatory variables and is easy to understand. The exact change for a 1% increase in explanatory variable would be β_1*log(1.01). In this case, however, absolute difference is very low and not of importance for conclusions drawn.

in two subsets and running regression on each of the subset individually is meaningful. The difference might also suggest that model 1 suffered from ecological fallacy – the conclusions derived from that were biased because there were two subsets of units, where each of them behaved differently.

Table 8 - Parameter estimates and other results of models 3M and 3F

	Model 3F			**Model 3M**		
	Estimate	SE	p-value	Estimate	SE	p-value
GDP	0.042	0.044	0.3449	0.138	0.030	<.0001
hlthexp	0.425	0.246	0.0860	0.678	0.148	<.0001
povrisk65	-0.058	0.031	0.0607	-0.062	0.032	0.0582
educ_high	-0.035	0.063	0.5835	-0.017	0.044	0.7025
educ_low	-0.138	0.059	0.0209	-0.120	0.046	0.0095
unem_long	0.044	0.070	0.5371	0.098	0.045	0.0298
R-squared	0.924			0.929		

Source: output of regression analysis performed in PANEL procedure in SAS 9.4

Generally, the model seems to explain variation in HLE with higher level of precision in new EU countries compared to the old ones. R-squared resulting from model 3 are 0.924 and 0.929, while in model 4, the values are 0.964 and 0.970. This implies that in old EU countries, there might be some other relatively important factors that influence HLE and that are omitted from the models.

Looking at models 3F and 3M more closely, only *educ_low* is significant for both sexes, suggesting that increase of one percentage point in proportion of population with low education is associated with a decrease in HLE by about 0.12 (for men) and 0.14 (for women) years. In model 1, this variable was significant only for men, and had lower magnitude.

In 3F, *educ_low* is at the same time the only significant variable. In 3M, estimates are significant also for *GDP*, *hlthexp* and *unem_long*. Parameters of *GDP* and *hlthexp* have expected positive directions, but the slope coefficient on *unem_long* is positive too, contrary to expectations.

Table 9 - Parameter estimates and other results of models 4M and 4F

	Model 4F			**Model 4M**		
	Estimate	SE	p-value	Estimate	SE	p-value
GDP	0.038	0.071	0.5946	0.247	0.079	0.0023
hlthexp	0.119	0.246	0.6283	0.501	0.281	0.0774
povrisk65	-0.040	0.018	0.0255	-0.038	0.018	0.0395
educ_high	0.387	0.047	<.0001	0.449	0.086	<.0001
educ_low	0.001	0.031	0.9665	0.048	0.051	0.3469
unem_long	-0.065	0.059	0.2702	0.126	0.049	0.0121
R-squared	0.964			0.970		

Source: output of regression analysis performed in PANEL procedure in SAS 9.4

In model 4 for new EU countries, estimates for *povrisk65* and *educ_high* are significant across both sexes. All estimates for these two variables are in the expected direction, and for *educ_high*, their absolute values is for both sexes also quite high. Compared to model 1, the estimates are higher by a factor of about two for women, and of about four for men. Conclusions concerning estimates on *GDP*, *hlthexp* and *unem_long* are almost the same in model 4 as in model 3. For women, all of these three are insignificant. For men, *GDP* and *unem_long* are significant with positive parameter estimates; *hlthexp* is estimated to have positive parameter as well, but the estimate is marginally insignificant (p-value = 0.077).

When comparing old and new EU countries on the basis of results of models 3 and 4, the estimates are consistent for GDP, long-term unemployment rate and partly for AROPE rate. *GDP* coefficient is significant for men only, and has positive direction. The same holds for *unem_long* – the consistently positive estimate for this variable across both groups of countries is against expectations and is commented on in second part of this chapter. *Povrisk65* is in all model specifications estimated to have negative coefficient, but conclusion about its significance differs between model 3 and model 4. In old EU countries, its p-values are slightly over the threshold of 0.05 (thus implying insignificance), while in new EU countries, they are significant, although not entirely persuasively (p-values being 0.026 and 0.040). This example show the arbitrariness of significance level – even though the difference in p-values seems not to be substantial, the decision about significance differs.

On the contrary, conclusion about association of HLE and education varies between the two groups. In old EU countries, proportion of population with low education is modelled to have significant relationship with HLE, but in new EU countries, the association seems to be between HLE and proportion of population with tertiary education. As for government expenditure on health, it seems to be positively associated with HLE, but significance was reached only for men in old EU countries.

Finally, models 5 and 6 follow the specification of model 2, but are disaggregated according to whether country belongs to the new members of EU or to the old members. The commentary for these two models is somewhat briefer, since many of the conclusions can be related to the models discussed previously.

Table 10 - Parameter estimates and other results of models 5M and 5F

	Model 5F			Model 5M		
	Estimate	SE	p-value	Estimate	SE	p-value
log(GDP)	2.878	1.735	0.0991	5.996	1.164	<.0001
hlthexp	0.485	0.239	0.0441	0.646	0.142	<.0001
povrisk65	-0.052	0.030	0.0866	-0.047	0.032	0.1412
educ_high	-0.048	0.061	0.4346	-0.006	0.039	0.8855
educ_low	-0.136	0.060	0.0239	-0.105	0.044	0.0172
unem_long	0.082	0.076	0.2837	0.136	0.048	0.0048
R-squared	0.925			0.931		

Source: output of regression analysis performed in PANEL procedure in SAS 9.4

Goodness-of-fit of models 5 and 6 is analogical to models 3 and 4, respectively – models analyzing new EU countries have somewhat higher R-squared than the ones for old members (Tables 10 and 11). Contrary to results of model 2, *log(GDP)* is in both models 5 and 6 significant only for men, not for women. Somewhat surprisingly, its coefficient is now in both groups of countries lower than previously, when all countries were taken together. With one percent rise in GDP, HLE is predicted to increase by 0.060 years in model 5M and 0.044 years in model 6M (Tables 10 and 11).

Table 11 - Parameter estimates and other results of models 6M and 6F

	Model 6F			**Model 6M**		
	Estimate	SE	p-value	Estimate	SE	p-value
log(GDP)	1.554	1.414	0.2739	4.380	1.339	0.0014
hlthexp	0.130	0.245	0.5963	0.477	0.282	0.0933
povrisk65	-0.038	0.017	0.0223	-0.026	0.019	0.1746
educ_high	0.365	0.051	<.0001	0.459	0.078	<.0001
educ_low	-0.002	0.031	0.9579	0.038	0.046	0.4099
unem_long	-0.039	0.062	0.5293	0.129	0.052	0.0139
R-squared	0.965			0.970		

Source: output of regression analysis performed in PANEL procedure in SAS 9.4

Expenditure on health is significant for both genders in model 5, although significance for women is not persuasive (p-value = 0.044). Parameter estimates of 0.485 in 5F and 0.646 in 5M are in line with previously reported estimates. Consistently with models 3 and 4, education is again estimated to be significantly associated with HLE in the form of *educ_high* in new member states, while in the old member states, the relationship is rather between HLE and *educ_low*. The magnitude of coefficients on variables concerning education tends to be very similar in models 5 and 6 as in models 3 and 4. Lastly, also results for long-term unemployment are well comparable to those reported previously. Coefficients are significant only for men and their values are positive.

7.2 Discussion of the results

In the previous section, results of the six models, where each of them was run separately for men and women, were addressed individually. This section attempts to generalize these results, sums up similarities and differences and matches the conclusions to theoretical assumptions presented in chapter 3 and to other studies with similar topic.

GDP per capita, as a first factor supposed to be associated with healthy life expectancy, was included in two different functional forms in the models – linear and logarithmic. Since the differences in goodness-of-fit were, regarding the functional form of GDP, very low, it is impossible to say which specification is better suited. However, the association of GDP and HLE is consistently significant only with men, not with women. Only one of the models suggested significant association for both sexes. This is not in line with results of Jagger et al.

(2008), who found strong relation between GDP and health. On the other hand, Fouweather et al. (2015) did not note any convincing relationship between these two in general. Results of the analysis in this thesis suggest that relationship between GDP and health actually exists, but that it might be mediated through other factors; that might be the reason for insignificance in most models for women. The general belief that health is positively linked to GDP is, however, not entirely supported by this analysis.

When looking at all countries together, the results for health expenditure are virtually the same as for GDP – significantly positive for men and insignificant for women. But unlike with GDP, government expenditure on health tends to behave differently in established EU countries and in new EU members. While in the established countries – that means mostly in Western, Northern and South-Western Europe – health expenditure tend to be significantly related to HLE also for women, this does not hold in the rest of the countries. Moreover, same change in the percentage of GDP spent on health is supposed to raise HLE by a greater amount in the established countries as opposed to the new ones.

An unclear conclusion about the association between health expenditure and health itself is generally in line with the assumptions. The significant linkage for men is more likely to support results of e.g. Crémieux et al. (1999), who also found such a relationship. On the other hand, the inconclusiveness suggested by Anderson and Frogner (2008) is in this thesis represented by the failure to find significant association for women.

Measures of poverty risk and education yield, out of the factors studied, most consistent results across both sexes and all model specifications. These two areas comprise three variables that were analyzed in regressions – AROPE rate for people aged 65 or more, proportion of population with low education and proportion of population with high education.

The AROPE rate is in all models estimated to have a negative association with HLE. This is a straightforward conclusion, but some caution is needed since parameters estimated for this variable are not significant in all cases. Significant negative relation was found in models evaluating all countries together, while the evidence was weaker in models 3 to 6. This, however, can be a consequence of smaller number of analyzed units and resulting higher standard errors in these latter models.

AROPE rate is also somewhat strongly correlated with GDP per capita (value of Pearson's correlation coefficient is equal to -0.61 for women's AROPE rate) and it is therefore possible that part of the relation between GDP and health is mediated by this measure of poverty risk. The pathways explaining the relationship with health are to some extent similar for both GDP and AROPE rate – they are based on accessibility of goods and services that influence health of an individual (e.g. Subramanian et al., 2002; Cutler et al., 2006; Jagger et al., 2008). The insignificance of coefficients on GDP in some specifications might therefore suggest that the link between health and wealth is in fact represented by proportion of people who are at risk of poverty or social exclusion, rather than by GDP itself.

As for education, there seem to be different relationships in the two country subgroups. In all countries together, HLE tends to be significantly associated with proportion of female population with high education and proportion of male population with low education. But models 3 to 6 reveal that in established EU countries, parameters are significant for *educ_low*, while in new EU countries, it is the other educational variable – *educ_high* – that is significantly linked to health. In both cases, values of parameters suggest that the higher level of education in society, the better health – this conclusion is in line with assumptions discussed in chapter 3.

Jagger et al. (2008) raised some questions in their work, since they found a positive relationship between HLY and proportion of population with low education in new EU member countries and concluded that the pathway between education and health is not straightforward. This thesis suggests different conclusions, but a remark of Jagger et al. (2008) in the sense that former communist countries have historically quite highly educated population, might be the reason for different results in new and established EU countries.

The last analyzed variable is long-term unemployment rate, where – if anything – positive link to healthy life expectancy was found. This link is significant only for men in most specifications, suggesting that countries with higher unemployment have higher HLE for men. This is not an expected result, and suggests that further aspects need to be considered in order to evaluate the true effect of unemployment on health (Béland et al., 2002).

Chapter 8

Conclusion

Length of human life is one of the central topics in demography, and has been studied by researchers for centuries. Length of healthy life, on the other hand, is a much younger concept, occurring in population studies only for a couple of decades. Despite that, health expectancies – measures of length of healthy life – are currently already well established in demographic research. Population (not only) in developed countries is ageing, which brings new challenges to governments, policy-makers, and in effect, to the society as a whole. Health expectancies bring valuable insights into debates about such challenges.

This thesis analyzed health expectancy in European countries between 2005 and 2017. Two objectives were set in the beginning – the first one was to describe situation and trends in health expectancy and its relation to life expectancy. The second one was to explain variation in health expectancy across European countries with the help of a set of measures from various fields. The first chapters were dedicated to theoretical discussion of measures of health expectancy, introduced history and presence of the indicators, as well as some computational and conceptual issues that are relevant for this topic. Before turning to analytical part of the thesis, some existing literature on the links of health to income, education, health spending and unemployment was presented in chapter 3. Chapter 4 presented data that were used for the analysis.

The first objective was primarily fulfilled with exploratory analysis performed and described in chapter 5. To start with, it is important to stress that conclusions about the trends were influenced greatly by the specific measure of health analyzed. Healthy life years, although being a structural indicator of Eurostat, suffered throughout the analyzed period from discrepances in definitions of underlying survey question across countries. Such discrepances made comparisons between countries and across time very complicated. Furthermore, a consolidated evaluation of the inter-country differences in wording of GALI was at the time of data collection available only until 2012. This finding raises some questions about validity of studies that only analyzed inter-country differences in HLY in one year (Jagger et al., 2008; Fouweather et al., 2015). The finding also had substantial connotations for further analysis in this thesis – healthy life expectancy based on self-perceived health was used as a primary measure in the remainder of the analyses – and, specifically, for evaluation of the hypotheses.

HLE was not influenced by changes in definitions, and enabled thus a proper evaluation of health trends. In general, countries of Northern, Western and Southern Europe had higher HLE throughout the period than countries in the Eastern part of the continent. This holds for absolute values of HLE as well as for its ratio to life expectancy. At the same time, growth of the values appeared to be faster in the countries with worse starting position, showing therefore some signs of convergence (which was, however, not examined formally). No substantial differences were found between sexes or with respect to whether it was HLE at birth, at age 50 or at age 65.

The second objective is fulfilled by evaluation of the hypotheses that were set in chapter 3. These are now repeated and one by one commented on, using the findings from chapter 7. Variation in HLE was explained with the help of GDP in purchasing power standard, share of GDP spent by government on health, at-risk-of-poverty or social exclusion rate for people aged above 65, long-term unemployment, proportion of population with low education (lower secondary at most) and share of population with tertiary education. There were 6 models formulated, which differed in functional form of GDP and in whether all countries, only new EU member countries or only established EU countries were included in the analysis.

- **Hypothesis 1**: *Gross domestic product per capita is positively associated with health expectancy.*

  Although this hypothesis was, according to the literature review, believed most strongly to hold, results show, to some extent, otherwise. GDP was shown to be consistently significantly associated with HLE only for men. The direction was positive, as was expected. A possible reason why the association for women was not significant is that the effect of wealth on health can be manifested through other variables, but not through GDP itself. Nevertheless, Hypothesis 1 **cannot be considered to hold** entirely.

- **Hypothesis 2**: *Income inequality is negatively associated with health expectancy.*

  Level of income inequality was in the regression models represented by AROPE rate for people aged above 65. The association was found in all models to be negative, but was significant only in those models that comprised all 31 countries. Since this relationship is supported by literature that lists a number of potential pathways between the two phenomena, Hypothesis 2 **is considered to hold**.

- **Hypothesis 3**: *Education has positive impact on health expectancy.*

  The positive linkage between education and health was proven by many researchers before (e.g. Ross and Wu, 1995; Cutler and Lleras-Muney, 2006; Conti et al., 2010) and is supported also by results of models in this thesis. A difference was found betwen the old and the new EU countries – while HLE appeared to be significantly related to proportion of population with low education in the old EU countries, new EU members showed significant association between HLE and proportion of tertiary-educated population. In both cases, however, higher education meant generally higher length of healthy life. Hypothesis 3 therefore, based on the results, **holds**.

- **Hypothesis 4**: *Unemployment rates are negatively associated with health expectancy.*

  It was anticipated that relationship between unemployment and health is complicated. Results did not provide a clear answer as for the direction or significance of the association – if anything, unemployment was shown to be positively associated with health for men. In view of that, Hypothesis 4 **is considered not to hold**.

To sum up the results found, it can be said that investments in increasing the education level of the population and in decreasing the inequality in society are most likely to have positive effect on health expectancy. Although causality cannot be inferred from the analysis directly, the association between HLE and education and inequality, respectively, is supported by theoretical pathways that tend to lead from education and inequality towards health. Education improves health due to increasing productivity (Grossman, 1976; Cutler and Lleras-Muney, 2006), higher chances for working in a safe environment (Cutler and Lleras-Muney, 2006) or lower tendency of educated people to engage in risky behaviour (Hunt-McCool and Bishop, 1998). Reducing inequality in society improves population health via better access of individuals to health care (Subramanian et al., 2002) or via an increase in social capital that coincides with lifting individuals out of poverty (Kawachi et al., 1997). The effect of the rest of the studied factors is speculative, and although some theoretical evidence is available for the relationship between unemployment and health, health spending and health levels, or GDP and health, results of analysis performed in this thesis do not provide clear conclusions.

While it is believed that this thesis is a worthy contribution towards the debate about health expectancies, it still has some limitations. Firstly, more advanced statistical models could be used, so that causality can be proved instead of a simple association. Secondly, some adjustments could be done to data on healthy life years, so that they are more comparable across time and countries and some conclusions can be drawn based on them. However, any such adjustment will demand detailed analysis of GALI question wording in all countries, since the latest available data that are provided by Eurostat on this topic are currently out-of-date. These issues thus do not have to be seen only as limitations, but rather as suggestions for what any future research can tackle.

REFERENCES

ALAVINIA, Seyed Mohammad and Alex BURDORF. 2008. Unemployment and retirement and ill-health: a cross-sectional analysis across European countries. *International Archives of Occupational and Environmental Health* [online]. 2008, **82**(1), 39-45 [cit. 2019-10-13]. DOI: 10.1007/s00420-008-0304-6. ISSN 0340-0131. Available at: http://link.springer.com/10.1007/s00420-008-0304-6

ANDERSON, Gerard F. and Bianca K. FROGNER. 2008. Health Spending In OECD Countries: Obtaining Value Per Dollar. *Health Affairs* [online]. 2008, **27**(6), 1718-1727 [cit. 2019-10-23]. DOI: 10.1377/hlthaff.27.6.1718. ISSN 0278-2715. Available at: http://www.healthaffairs.org/doi/10.1377/hlthaff.27.6.1718

BALTAGI, Badi H. *Econometric analysis of panel data*. 3rd edition. Hoboken, NJ: John Wiley, 2005. ISBN 978-0-470-01456-1.

BÉLAND, François, Stephen BIRCH and Greg STODDART. 2002. Unemployment and health: contextual-level influences on the production of health in populations. *Social Science & Medicine* [online]. 2002, **55**(11), 2033-2052 [cit. 2019-10-13]. DOI: 10.1016/S0277-9536(01)00344-6. ISSN 02779536. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0277953601003446

BLOOM, David E. and David CANNING. 2000. The Health and Wealth of Nations. *Science* [online]. 2000, **287**(5456), 1207-1209 [cit. 2019-10-13]. DOI: 10.1126/science.287.5456.1207. ISSN 00368075. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.287.5456.1207

BOGAERT, Petronille, Herman VAN OYEN, Isabelle BELUCHE, Emmanuelle CAMBOIS and Jean-Marie ROBINE. 2018. The use of the global activity limitation Indicator and healthy life years by member states and the European Commission. *Archives of Public Health* [online]. 2018, **76**(30) [cit. 2020-01-11]. DOI: 10.1186/s13690-018-0279-z. ISSN 2049-3258. Available at: https://archpublichealth.biomedcentral.com/articles/10.1186/s13690-018-0279-z

BONE, Margaret R. 1992. International Efforts to Measure Health Expectancy. *Journal of Epidemiology and Community Health (1979-)*. 1992, **46**(6), 555-558. Available also at: https://www.jstor.org/stable/25567629

BRØNNUM-HANSEN, Henrik, Mette Lindholm ERIKSEN, Karen ANDERSEN-RANBERG and Bernard JEUNE. 2017. Persistent social inequality in life expectancy and disability-free life expectancy: Outlook for a differential pension age in Denmark? *Scandinavian Journal of Public Health* [online]. 2017, **45**(4), 459-462 [cit. 2020-03-22]. DOI: 10.1177/1403494816683591. ISSN 1403-4948. Available at: http://journals.sagepub.com/doi/10.1177/1403494816683591

CAMBOIS, Emmanuelle, Aurore CLAVEL, Isabelle ROMIEU and Jean-Marie ROBINE. 2008. Trends in disability-free life expectancy at age 65 in France: consistent and diverging patterns according to the underlying disability measure. *European Journal of Ageing* [online]. 2008, **5**(4), 287-298 [cit. 2019-11-24]. DOI: 10.1007/s10433-008-0097-1. ISSN 1613-9372. Available at: http://link.springer.com/10.1007/s10433-008-0097-1

CONTI, Gabriella, James HECKMAN and Sergio URZUA. 2010. The Education-Health Gradient. *American Economic Review* [online]. 2010, **100**(2), 234-238 [cit. 2019-10-26]. DOI: 10.1257/aer.100.2.234. ISSN 0002-8282. Available at: https://pdfs.semanticscholar.org/d91e/167f3eaa57d27148ecd4f5d66a4d6e091509.pdf?_ga=2.168830564.418209809.1578131036-1427978451.1578131036

CRÉMIEUX, Pierre-Yves, Pierre OUELLETTE and Caroline PILON. 1999. Health care spending as determinants of health outcomes. *Health Economics* [online]. 1999, **8**(7), 627-639 [cit. 2019-10-13]. DOI: 10.1002/(SICI)1099-1050(199911)8:7<627::AID-HEC474>3.0.CO;2-8. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291099-1050%28199911%298%3A7%3C627%3A%3AAID-HEC474%3E3.0.CO%3B2-8

CUTLER, David, Angus DEATON and Adriana LLERAS-MUNEY. 2006. The Determinants of Mortality. *Journal of Economic Perspectives* [online]. 2006, **20**(3), 97-120 [cit. 2019-10-13]. DOI: 10.1257/jep.20.3.97. ISSN 0895-3309. Available at: http://pubs.aeaweb.org/doi/10.1257/jep.20.3.97

CUTLER, David M. and Adriana LLERAS-MUNEY. 2006. Education and Health: Evaluating Theories and Evidence. *NBER Working Papers* [online]. 2006, **12352** [cit. 2019-10-26]. DOI: 10.3386/w12352. Available at: https://www.nber.org/papers/w12352

EUROPEAN COMMISSION. 2011. Manual on sources and methods for the compilation of COFOG statistics — Classification of the Functions of Government (COFOG). *Methodologies & Working papers* [online]. Luxembourg, 2011 [cit. 2019-12-14]. DOI: 10.2785/16355. Available at: https://ec.europa.eu/eurostat/documents/3859598/5917333/KS-RA-11-013-EN.PDF/2eb9714a-ee4b-49fe-baab-e9af5ca457b1

EUROSTAT. 2012. Healthy life expectancy based on self-perceived health. *Eurostat* [online]. Luxembourg, 2012 [cit. 2019-09-07]. Available at: https://ec.europa.eu/eurostat/cache/metadata/en/hlth_silc_17_esms.htm

EUROSTAT. 2014a. Healthy life years (from 2004 onwards). *Eurostat* [online]. Luxembourg, 2014 [cit. 2019-09-07]. Available at: https://ec.europa.eu/eurostat/cache/metadata/en/hlth_hlye_esms.htm

EUROSTAT. 2014b. *Working paper with the description of the 'Income and living conditions dataset'* [online]. Luxembourg: Eurostat, 2014 [cit. 2019-11-24]. Available at: https://ec.europa.eu/eurostat/documents/1012329/1012398/D5.1.3-Working_paper_final_20141204.pdf/c4ed99f5-7cc3-4bf9-a6eb-b1f730299e0e

EUROSTAT. 2017a. Mortality. *Eurostat* [online]. Luxembourg, 2017 [cit. 2019-11-23]. Available at: https://ec.europa.eu/eurostat/cache/metadata/en/demo_mor_esms.htm

EUROSTAT. 2017b. Income and living conditions. *Eurostat* [online]. Luxembourg, 2017 [cit. 2019-11-24]. Available at: https://ec.europa.eu/eurostat/cache/metadata/en/ilc_esms.htm

EUROSTAT. 2017c. Educational attainment level and transition from education to work (based on EU-LFS). *Eurostat* [online]. Luxembourg, 2017 [cit. 2019-11-24]. Available at: https://ec.europa.eu/eurostat/cache/metadata/en/edat1_esms.htm

EUROSTAT. 2018. Unemployment - LFS adjusted series. *Eurostat* [online]. Luxembourg, 2018 [cit. 2019-11-24]. Available at: https://ec.europa.eu/eurostat/cache/metadata/en/une_esms.htm

EUROSTAT. 2019a. Annual national accounts. *Eurostat* [online]. Luxembourg, 2019 [cit. 2019-11-24]. Available at: https://ec.europa.eu/eurostat/cache/metadata/en/nama10_esms.htm

EUROSTAT. 2019b. EU labour force survey - methodology. *Eurostat* [online]. Luxembourg, 2019 [cit. 2019-11-24]. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/EU_labour_force_survey_-_methodology

EUROSTAT. 2019c. General government expenditure by function (COFOG). *Eurostat* [online]. Luxembourg, 2019 [cit. 2019-12-14]. Available at: https://ec.europa.eu/eurostat/cache/metadata/en/gov_10a_exp_esms.htm

EUROSTAT. 2019d. TF SILC Revision: Point 9. Breaks in series. *Eurostat* [online]. Luxembourg, 2019 [cit. 2020-01-11]. Available at: https://circabc.europa.eu/webdav/CircaBC/ESTAT/eusilc/Library/wgtf/revision_eu-silc/18th%20meeting%2014th%20May%202019/LC-LEGAL-06-19%20Break%20in%20Series.pdf

FOUWEATHER, Tony, Clare GILLIES, Pia WOHLAND, Herman VAN OYEN, Wilma NUSSELDER, Jean-Marie ROBINE, Emmanuelle CAMBOIS and Carol JAGGER. 2015. Comparison of socio-economic indicators explaining inequalities in Healthy Life Years at age 50 in Europe: 2005 and 2010. *The European Journal of Public Health* [online]. 2015, **25**(6), 978-983 [cit. 2019-11-24]. DOI: 10.1093/eurpub/ckv070. ISSN 1101-1262. Available at: https://academic.oup.com/eurpub/article-lookup/doi/10.1093/eurpub/ckv070

FRIES, James F. 1980. Aging, Natural Death, and the Compression of Morbidity. *New England Journal of Medicine* [online]. 1980, **303**(3), 130-135 [cit. 2019-09-08]. DOI: 10.1056/NEJM198007173030304. ISSN 0028-4793. Available at: https://europepmc.org/abstract/med/7383070

GROSSMAN, Michael. 1976. The Correlation between Health and Schooling. *Household production and consumption* [online]. National Bureau of Economic Research: Columbia University Press, 1976, s. 147-224 [cit. 2019-10-26]. ISBN 0-870-14515-0. Available at: https://www.nber.org/chapters/c3962

GRUENBERG, Ernest M. 1977. The Failures of Success. *The Milbank Memorial Fund Quarterly. Health and Society* [online]. 1977, **55**(1) [cit. 2019-09-08]. DOI: 10.2307/3349592. ISSN 01601997. Available at: https://www.jstor.org/stable/3349592

GUPTA, Sanjeev, Marijn VERHOEVEN and Erwin R TIONGSON. 2002. The effectiveness of government spending on education and health care in developing and transition economies. *European Journal of Political Economy* [online]. 2002, **18**(4), 717-737 [cit. 2019-10-26]. DOI: 10.1016/S0176-2680(02)00116-7. ISSN 01762680. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0176268002001167

GUTIERREZ-FISAC, Juan L., Rosa GISPERT and Judit SOLA. 2000. Factors explaining the geographical differences in Disability Free Life Expectancy in Spain. *Journal of Epidemiology and Community Health* [online]. 2000, **54**(6), 451-455 [cit. 2019-10-13]. DOI: 10.1136/jech.54.6.451. ISSN 0143005X. Available at: http://jech.bmj.com/cgi/doi/10.1136/jech.54.6.451

HOLUB, Martin, Štěpánka POLLNEROVÁ and Milan ŠLAPÁK. 2015. *Způsoby automatického přizpůsobení důchodového věku naději dožití v zemích EU* [online]. Praha: VÚPSV, 2015 [cit. 2020-03-22]. ISBN 978-80-7416-220-6.

HOWSE, Kenneth. 2006. Increasing Life Expectancy and the Compression of Morbidity: A Critical Review of the Debate. In: *Oxford Institute of Ageing Working Papers* [online]. Oxford, 2006 [cit. 2019-09-14]. Available at: https://www.researchgate.net/publication/254659502

HUNT-MCCOOL, Janet and Dawn M BISHOP. 1998. Health economics and the economics of education: specialization and division of labor. *Economics of Education Review* [online].

1998, **17**(3), 237-244 [cit. 2019-10-26]. DOI: 10.1016/S0272-7757(97)00045-9. ISSN 02727757. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0272775797000459

IMAI, Kosuke and Samir SONEJI. 2007. On the Estimation of Disability-Free Life Expectancy. *Journal of the American Statistical Association* [online]. 2007, **102**(480), 1199-1211 [cit. 2019-09-14]. DOI: 10.1198/016214507000000040. ISSN 0162-1459. Available at: http://www.tandfonline.com/doi/abs/10.1198/016214507000000040

INED. 2019. About REVES. *REVES* [online]. INED, 2019 [cit. 2019-09-08]. Available at: https://reves.site.ined.fr/en/home/about_reves/

JAGGER, Carol, Clare GILLIES, Emmanuelle CAMBOIS, Herman VAN OYEN, Wilma NUSSELDER and Jean-Marie ROBINE. 2010. The Global Activity Limitation Index measured function and disability similarly across European countries. *Journal of Clinical Epidemiology* [online]. 2010, **63**(8), 892-899 [cit. 2019-11-24]. DOI: 10.1016/j.jclinepi.2009.11.002. ISSN 08954356. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0895435609003485

JAGGER, Carol, Clare GILLIES, Francesco MOSCONE, Emmanuelle CAMBOIS, Herman VAN OYEN, Wilma NUSSELDER and Jean-Marie ROBINE. 2008. Inequalities in healthy life years in the 25 countries of the European Union in 2005: a cross-national meta-regression analysis. *The Lancet* [online]. 2008, **372**(9656), 2124-2131 [cit. 2019-10-12]. DOI: 10.1016/S0140-6736(08)61594-9. ISSN 01406736. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0140673608615949

JAGGER, Carol, Martin MCKEE, Kaare CHRISTENSEN, et al. 2013. Mind the gap—reaching the European target of a 2-year increase in healthy life years in the next decade. *European Journal of Public Health* [online]. 2013, **23**(5), 829-833 [cit. 2020-01-11]. DOI: 10.1093/eurpub/ckt030. ISSN 1464-360X. Available at: https://academic.oup.com/eurpub/article-lookup/doi/10.1093/eurpub/ckt030

JAGGER, Carol and Jean-Marie ROBINE. 2011. Healthy Life Expectancy. *International Handbook of Adult Mortality* [online]. Dordrecht: Springer Netherlands, 2011, 2011-1-5, , 551-568 [cit. 2019-09-14]. International Handbooks of Population. DOI: 10.1007/978-90-481-9996-9_26. ISBN 978-90-481-9995-2. Available at: http://www.springerlink.com/index/10.1007/978-90-481-9996-9_26

KABIR, Mahfuz. 2008. Determinants of Life Expectancy in Developing Countries. *The Journal of Developing Areas* [online]. 2008, **41**(2), 185-204 [cit. 2020-02-16]. Available at: www.jstor.org/stable/40376184

KALOUSOVA, Lucie and Sarah A. BURGARD. 2014. Unemployment, measured and perceived decline of economic resources: Contrasting three measures of recessionary hardships and their implications for adopting negative health behaviors. *Social Science & Medicine* [online]. 2014, **106**, 28-34 [cit. 2019-10-13].

DOI: 10.1016/j.socscimed.2014.01.007. ISSN 02779536. Available at: https://linkinghub.elsevier.com/retrieve/pii/S027795361400032X

KASSEBAUM, Nicholas J, Megha ARORA, Ryan M BARBER, et al. 2016. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015:: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*. 2016, **388**(10053), 1603-1658. DOI: 10.1016/S0140-6736(16)31460-X. ISSN 01406736. Available at: https://linkinghub.elsevier.com/retrieve/pii/S014067361631460X

KATZ, Sidney, Laurence G. BRANCH, Michael H. BRANSON, Joseph A. PAPSIDERO, John C. BECK and David S. GREER. 1983. Active Life Expectancy. *New England Journal of Medicine* [online]. 1983, **309**(20), 1218-1224 [cit. 2019-09-15]. DOI: 10.1056/NEJM198311173092005. ISSN 0028-4793. Available at: http://www.nejm.org/doi/abs/10.1056/NEJM198311173092005

KAWACHI, Ichiro, Bruce P KENNEDY, Kimberly LOCHNER and Deborah PROTHROW-STITH. 1997. Social capital, income inequality, and mortality. *American Journal of Public Health* [online]. 1997, **87**(9), 1491-1498 [cit. 2019-10-12]. DOI: 10.2105/AJPH.87.9.1491. ISSN 0090-0036. Available at: http://ajph.aphapublications.org/doi/10.2105/AJPH.87.9.1491

LADITKA, James N. and Sarah B. LADITKA. 2016. Unemployment, disability and life expectancy in the United States: A life course study. *Disability and Health Journal* [online]. 2016, **9**(1), 46-53 [cit. 2019-10-13]. DOI: 10.1016/j.dhjo.2015.08.003. ISSN 19366574. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1936657415001077

LESTHAEGHE, Ron and Dirk VAN DE KAA. 1986. Twee demografische transities. In: VAN DE KAA, Dirk J and Ron LESTHAEGHE. *Bevolking: groei en krimp* [online]. Deventer: Van Loghum Slaterus, 1986, s. 9-24 [cit. 2020-03-22].

MAJER, I. M., W. J. NUSSELDER, J. P. MACKENBACH and A. E. KUNST. 2011. *Socioeconomic inequalities in life and health expectancies around official retirement age in 10 Western-European countries* [online]. 2011, **65**(11), 972-979 [cit. 2020-03-22]. DOI: 10.1136/jech.2010.111492. ISSN 0143-005X. Available at: http://jech.bmj.com/cgi/doi/10.1136/jech.2010.111492

MANKIW, N. Gregory. 2012. *Principles of Economics*. 6th edition. Mason, USA: South-Western Cengage Learning, 2012. ISBN 978-0-538-45305-9.

MANTON, Kenneth G. 1982. Changing Concepts of Morbidity and Mortality in the Elderly Population. *The Milbank Memorial Fund Quarterly. Health and Society* [online]. 1982, **60**(2), 183-244 [cit. 2019-09-14]. DOI: 10.2307/3349767. ISSN 0160-1997. Available at: https://www.jstor.org/stable/3349767?origin=crossref

MATHERS, C. D., J.-M. ROBINE and R. WILKINS. 1994. Health expectancy indicators: recommendations for terminology. *Advances in health expectancies*. Canberra: Australian Institute on Health and Welfare, 1994, s. 34-41.

MATHERS, Colin D. 2002. Health Expectancies: an Overview and Critical Appraisal. *Summary measures of population health: concepts, ethics, measurement, and applications*. Geneva: World Health Organization, 2002, s. 177-204. ISBN 9241545518.

MATHERS, Colin D and Jean-Marie ROBINE. 1997. *How good is Sullivan's method for monitoring changes in population health expectancies?* [online]. 1997, **51**(1), 80-86 [cit. 2019-09-14]. DOI: 10.1136/jech.51.1.80. ISSN 0143-005X. Available at: http://jech.bmj.com/cgi/doi/10.1136/jech.51.1.80

MATHERS, Colin D, Ritu SADANA, Joshua A SALOMON, Christopher JL MURRAY and Alan D LOPEZ. 2001. Healthy life expectancy in 191 countries, 1999. *The Lancet*. 2001, **357**(9269), 1685-1691. DOI: 10.1016/S0140-6736(00)04824-8. ISSN 01406736. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0140673600048248

MATTHEWS, Ruth J., Lucy K. SMITH, Ruth M. HANCOCK, Carol JAGGER and Nicola A. SPIERS. 2005. *Socioeconomic factors associated with the onset of disability in older age: a longitudinal study of people aged 75 years and over* [online]. 2005, **61**(7), 1567-1575 [cit. 2019-10-12]. DOI: 10.1016/j.socscimed.2005.02.007. ISSN 02779536. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0277953605000845

MINAGAWA, Yuka. 2013. Inequalities in Healthy Life Expectancy in Eastern Europe. *Population and Development Review* [online]. 2013, **39**(4), 649-671 [cit. 2020-02-15]. DOI: 10.1111/j.1728-4457.2013.00632.x. ISSN 00987921. Available at: http://doi.wiley.com/10.1111/j.1728-4457.2013.00632.x

MURRAY, Christopher J L, Ryan M BARBER, Kyle J FOREMAN, et al. 2015. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying the epidemiological transition. *The Lancet*. 2015, **386**(10009), 2145-2191. DOI: 10.1016/S0140-6736(15)61340-X. ISSN 01406736. Available at: https://linkinghub.elsevier.com/retrieve/pii/S014067361561340X

NEWEY, Whitney K. and Kenneth D. WEST. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* [online]. 1987, **55**(3), 703-708 [cit. 2020-03-07]. DOI: 10.2307/1913610. ISSN 00129682. Do Available at: https://www.jstor.org/stable/1913610?origin=crossref

PARK, Hun Myoung. 2011. Practical Guides To Panel Data Modeling: A Step by Step Analysis Using Stata. *Tutorial Working Paper* [online]. Graduate School of International Relations, International University of Japan, 2011 [cit. 2020-02-15]. Available at: https://www.iuj.ac.jp/faculty/kucc625/method/panel/panel_iuj.pdf

PARK, M. Jane, Claire D. BRINDIS, Fay CHANG and Charles E. IRWIN. 2008. A Midcourse Review of the Healthy People 2010: 21 Critical Health Objectives for Adolescents and Young Adults. *Journal of Adolescent Health* [online]. 2008, **42**(4), 329-334 [cit. 2020-03-22]. DOI: 10.1016/j.jadohealth.2008.01.008. ISSN 1054139X. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1054139X08000827

PERENBOOM, RJM, LM VAN HERTEN, HC BOSHUIZEN and GAM VAN DEN BOS. 2016. Life Expectancy without Chronic Morbidity: Trends in Gender and Socioeconomic Disparities. *Public Health Reports* [online]. 2016, **120**(1), 46-54 [cit. 2019-09-15]. DOI: 10.1177/003335490512000109. ISSN 0033-3549. Available at: http://journals.sagepub.com/doi/10.1177/003335490512000109

RITCHIE, K, J M ROBINE, L LETENNEUR and J F DARTIGUES. 1994. Dementia-free life expectancy in France. *American Journal of Public Health* [online]. 1994, **84**(2), 232-236 [cit. 2019-09-15]. DOI: 10.2105/AJPH.84.2.232. ISSN 0090-0036. Available at: http://ajph.aphapublications.org/doi/10.2105/AJPH.84.2.232

ROBINE, Jean-Marie and Emannuelle CAMBOIS. 2013. Healthy life expectancy in Europe. *Population & Societies* [online]. 2013, **499**, 1-4 [cit. 2020-01-11]. ISSN 0184-7783. Available at: https://search.proquest.com/docview/1417586815?

ROBINE, Jean-Marie and Carol JAGGER. 2017. Health expectancies. MICHEL, Jean-Pierre, B. Lynn BEATTIE, Finbarr C. MARTIN a Jeremy WALSTON. *Oxford Textbook of Geriatric Medicine* [online]. 3rd. Oxford: Oxford University Press, 2017, s. 63-68 [cit. 2019-09-15]. ISBN 0191062626. Available at: https://books.google.cz/books?id=RAJCDwAAQBAJ

ROELFS, David J., Eran SHOR, Karina W. DAVIDSON and Joseph E. SCHWARTZ. 2011. Losing life and livelihood: A systematic review and meta-analysis of unemployment and all-cause mortality. *Social Science & Medicine* [online]. 2011, **72**(6), 840-854 [cit. 2019-10-13]. DOI: 10.1016/j.socscimed.2011.01.005. ISSN 02779536. Available at: https://linkinghub.elsevier.com/retrieve/pii/S027795361100044X

ROGERS, Andrei, Richard G ROGERS and Laurence G BRANCH. 1989. A Multistate Analysis of Active Life Expectancy. *Public Health Reports* [online]. 1989, **104**(3), 222-226 [cit. 2019-09-14]. Available at: https://www.jstor.org/stable/4628648

ROSS, Catherine E. and Chia-ling WU. 1995. The Links Between Education and Health. *American Sociological Review* [online]. 1995, **60**(5), 719-745 [cit. 2019-10-26]. DOI: 10.2307/2096319. ISSN 00031224. Available at: http://www.jstor.org/stable/2096319?origin=crossref

RYCHTAŘÍKOVÁ, Jitka. 2006. Zdravá délka života v současné české populaci. *Demografie* [online]. 2006, **48**, 166-178 [cit. 2020-04-10]. ISSN 0011-8265.

Available at: https://www.czso.cz/documents/10180/20563533/180306q3.pdf/d9512b5a-1cd8-423d-9d10-9fe6ee9da46d?version=1.0

SALOMON, Joshua A, Haidong WANG, Michael K FREEMAN, Theo VOS, Abraham D FLAXMAN, Alan D LOPEZ and Christopher JL MURRAY. 2012. Healthy life expectancy for 187 countries, 1990–2010: a systematic analysis for the Global Burden Disease Study 2010. *The Lancet*. 2012, **380**(9859), 2144-2162. DOI: 10.1016/S0140-6736(12)61690-0. ISSN 01406736. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0140673612616900

SANDERS, Barkev S. 1964. Measuring Community Health Levels. *American Journal of Public Health and the Nations Health*. 1964, **54**(7), 1063-1070. DOI: 10.2105/AJPH.54.7.1063. ISSN 0002-9572. Available at: http://ajph.aphapublications.org/doi/10.2105/AJPH.54.7.1063

SAS INSTITUTE. 2014. The PANEL Procedure. *SAS/ETS® 13.2 User's Guide* [online]. Cary, USA: SAS Institute, 2014, s. 1375-1493 [cit. 2020-03-29]. Available at: https://support.sas.com/documentation/onlinedoc/ets/132/panel.pdf

SUBRAMANIAN, S. V., Paolo BELLI and Ichiro KAWACHI. 2002. The Macroeconomic Determinants of Health. *Annual Review of Public Health* [online]. 2002, **23**(1), 287-302 [cit. 2019-10-13]. DOI: 10.1146/annurev.publhealth.23.100901.140540. ISSN 0163-7525. Available at: http://www.annualreviews.org/doi/10.1146/annurev.publhealth.23.100901.140540

SULLIVAN, Daniel F. 1971. A Single Index of Mortality and Morbidity. *HSMHA Health Reports*. 1971, **86**(4), 347-354. DOI: 10.2307/4594169. ISSN 0083-1204. Available at: https://www.jstor.org/stable/10.2307/4594169?origin=crossref

SULLIVAN, Daniel F. 1966. Conceptual problems in developing an index of health. *Public Health Service Publication Series No. 1000, Series 2, No. 17*. Washington DC, 1966.

VAN OYEN, Herman, Wilma NUSSELDER, Carol JAGGER, Petra KOLIP, Emmanuelle CAMBOIS and Jean-Marie ROBINE. 2013. Gender differences in healthy life years within the EU: an exploration of the "health–survival" paradox. *International Journal of Public Health* [online]. 2013, **58**(1), 143-155 [cit. 2020-01-11]. DOI: 10.1007/s00038-012-0361-1. ISSN 1661-8556. Available at: http://link.springer.com/10.1007/s00038-012-0361-1

VOS, Theo, Abraham D FLAXMAN, Mohsen NAGHAVI, et al. 2012. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010:: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2012, **380**(9859), 2163-2196. DOI: 10.1016/S0140-6736(12)61729-2. ISSN 01406736. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0140673612617292

WALBURN, David. Europe 2020. 2010. *Local Economy: The Journal of the Local Economy Policy Unit* [online]. 2010, **25**(8), 699-702 [cit. 2020-04-10].

DOI: 10.1080/02690942.2010.533420. ISSN 0269-0942. Available at: http://journals.sagepub.com/doi/10.1080/02690942.2010.533420

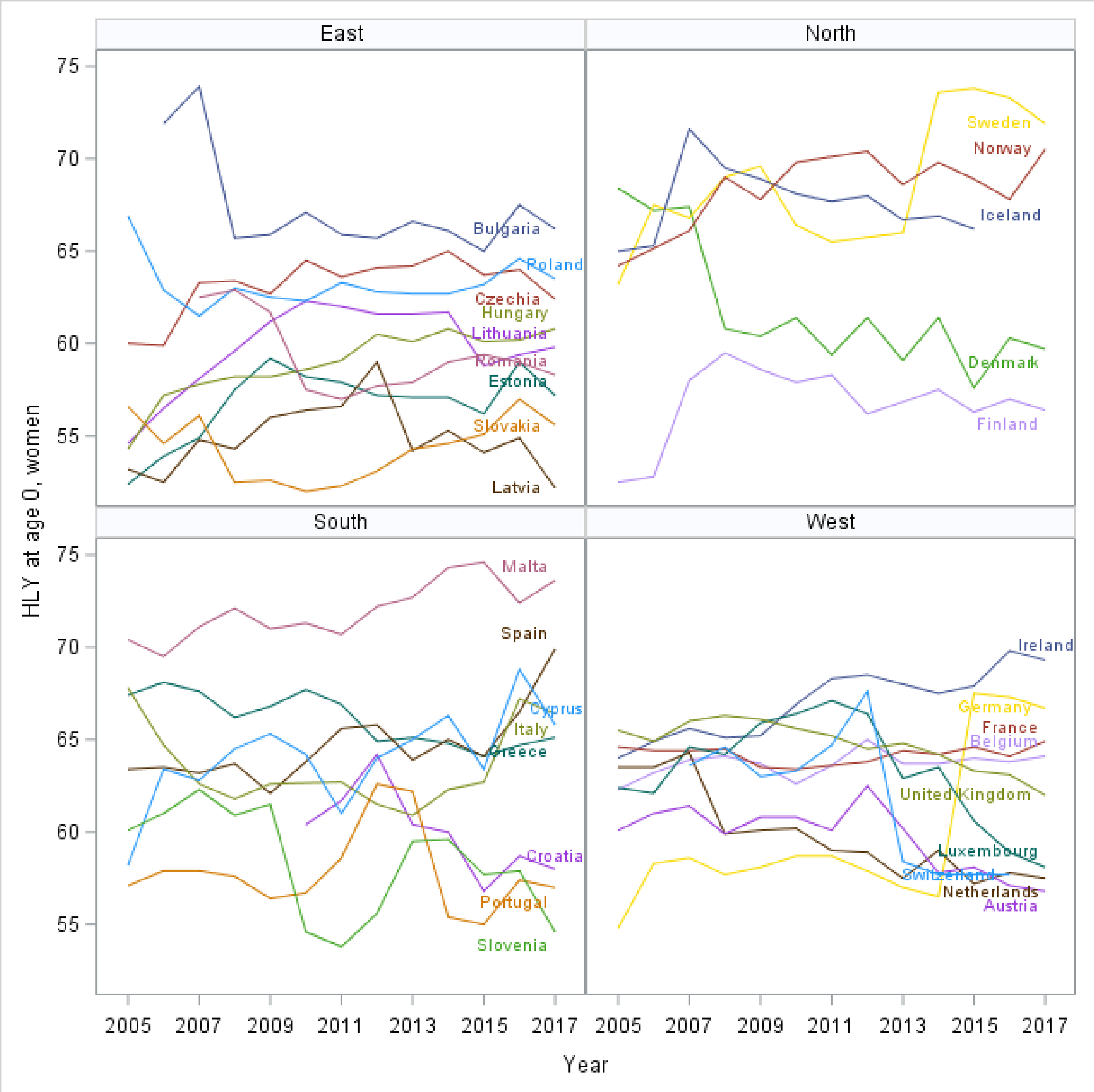
WILKINSON, Richard G., Ichiro KAWACHI and Bruce P. KENNEDY. 1998. Mortality, the Social Environment, Crime and Violence. *Sociology of Health* [online]. 1998, **20**(5), 578-597 [cit. 2019-10-12]. DOI: 10.1111/1467-9566.00120. ISSN 0141-9889. Available at: http://doi.wiley.com/10.1111/1467-9566.00120

WOLFE, Barbara L. 1985. The Influence of Health on School Outcomes: A Multivariate Approach. *Medical Care* [online]. 1985, **23**(10), 1127-1138 [cit. 2019-10-26]. Available at: https://www.jstor.org/stable/3764861

WOOLDRIDGE, Jeffrey M. 2018. *Introductory econometrics: a modern approach*. 7th edition. Mason, OH: Cengage Learning, 2018. ISBN 978-1-337-55886-0.

APPENDIX

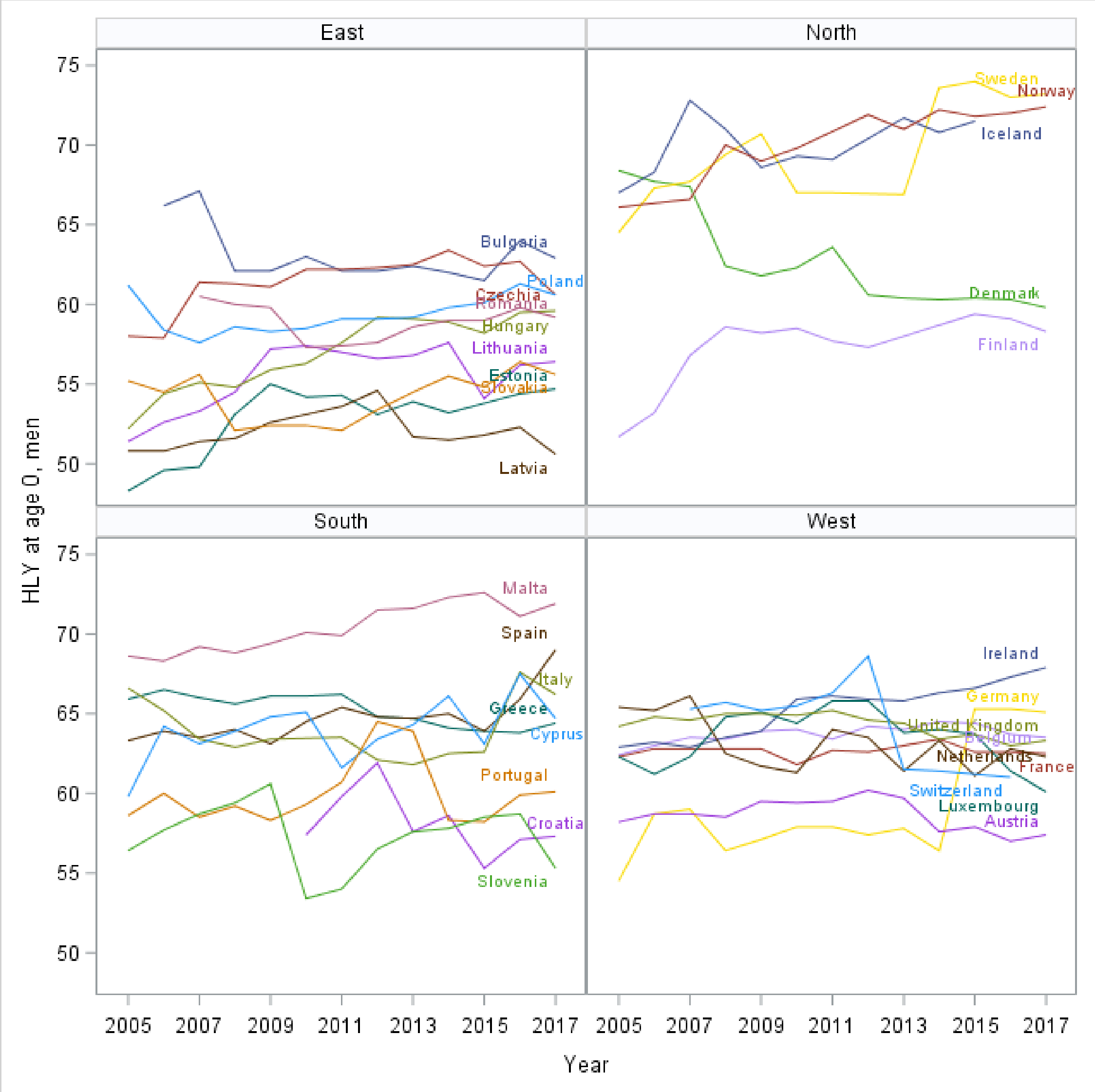
Appendix 1 – Healthy life years at age 0 in European countries, women, 2005–2017



Source: Eurostat

Note: processed by SAS 9.4

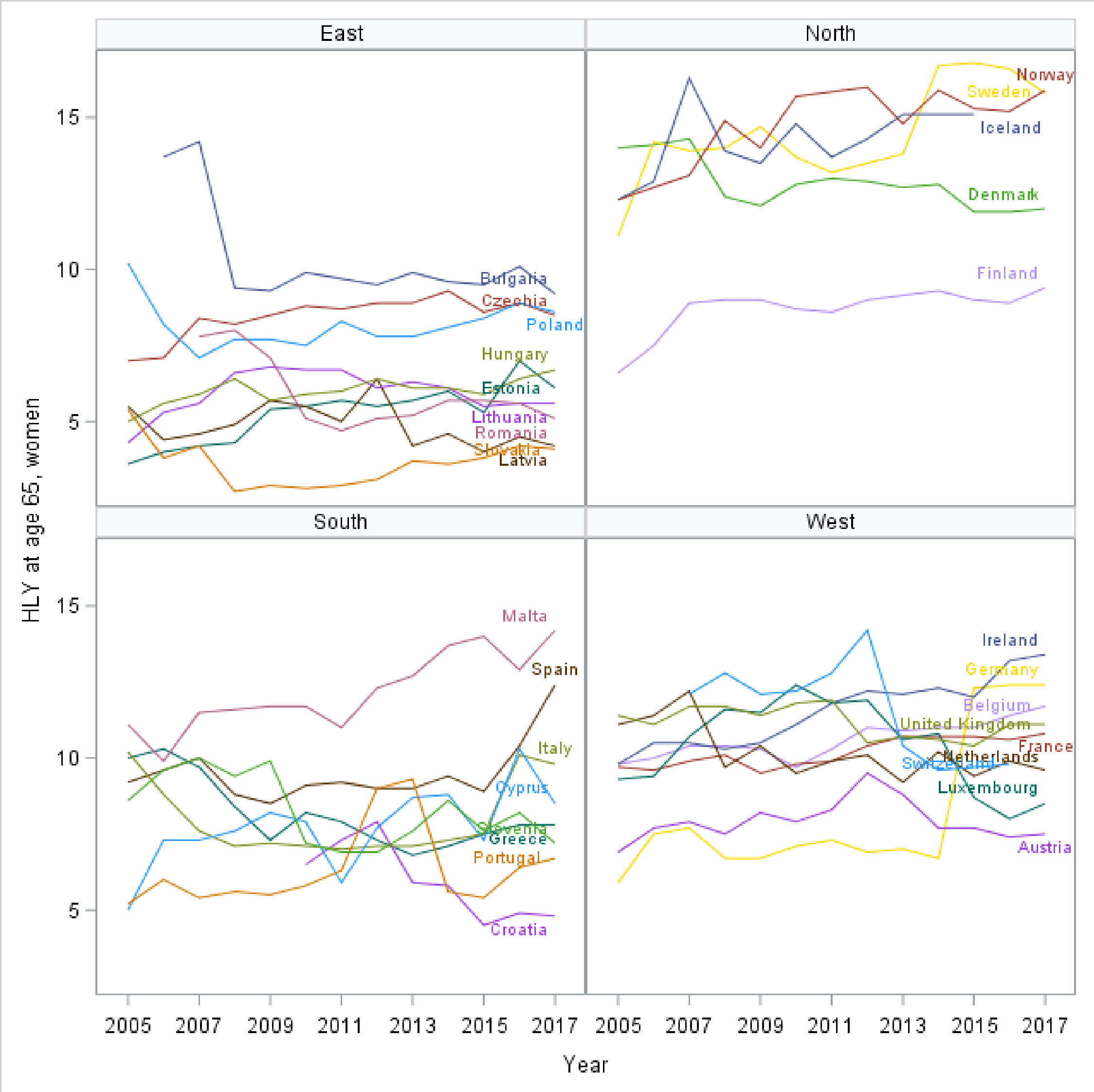
Appendix 2 – Healthy life years at age 0 in European countries, men, 2005–2017



Source: Eurostat

Note: processed by SAS 9.4

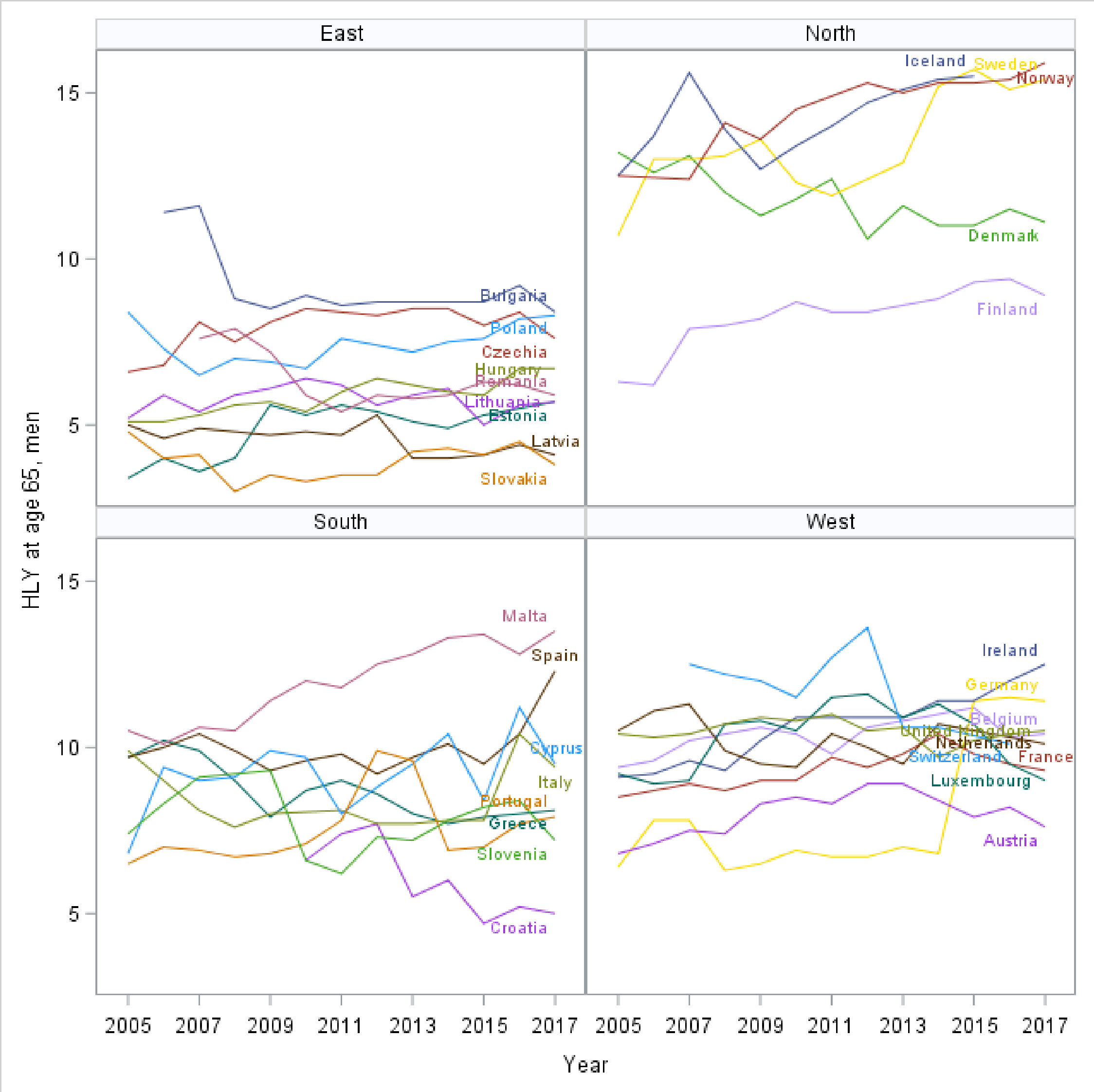
Appendix 3 – Healthy life years at age 65 in European countries, women, 2005–2017



Source: Eurostat

Note: processed by SAS 9.4

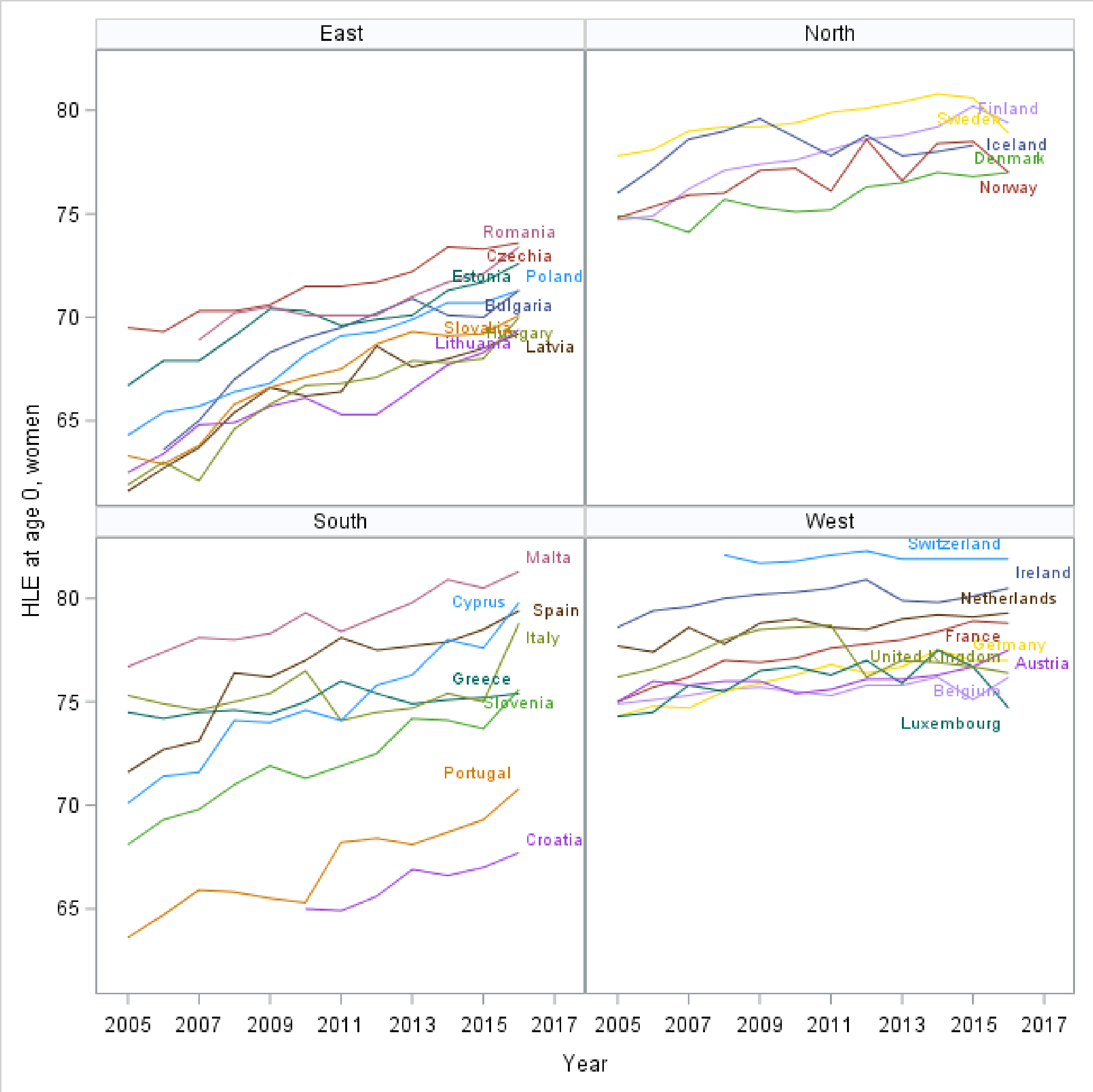
Appendix 4 – Healthy life years at age 65 in European countries, men, 2005–2017



Source: Eurostat

Note: processed by SAS 9.4

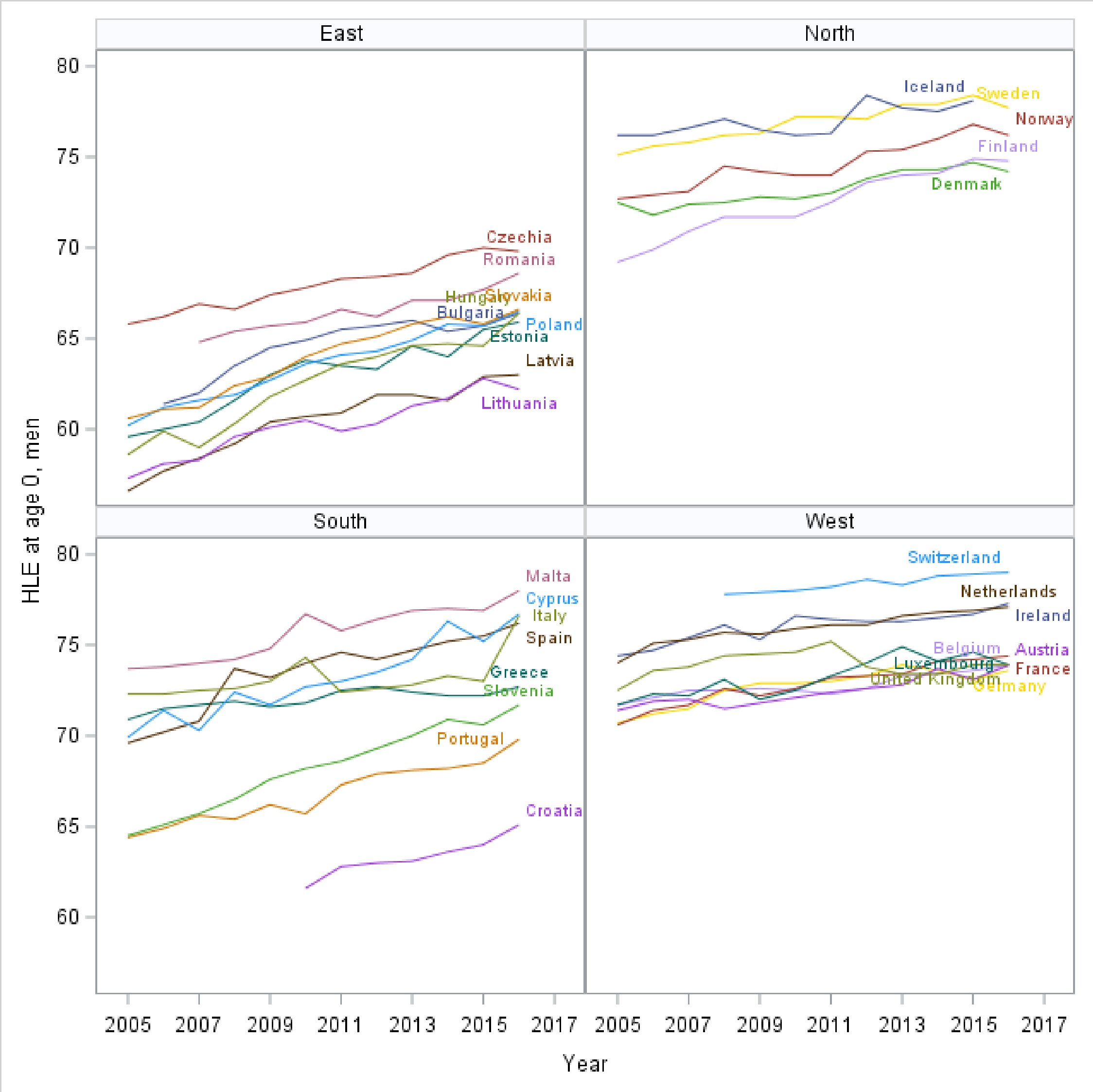
Appendix 5 – Healthy life expectancy based on self-perceived health at age 0 in European countries, women, 2005–2016



Source: Eurostat

Note: processed by SAS 9.4

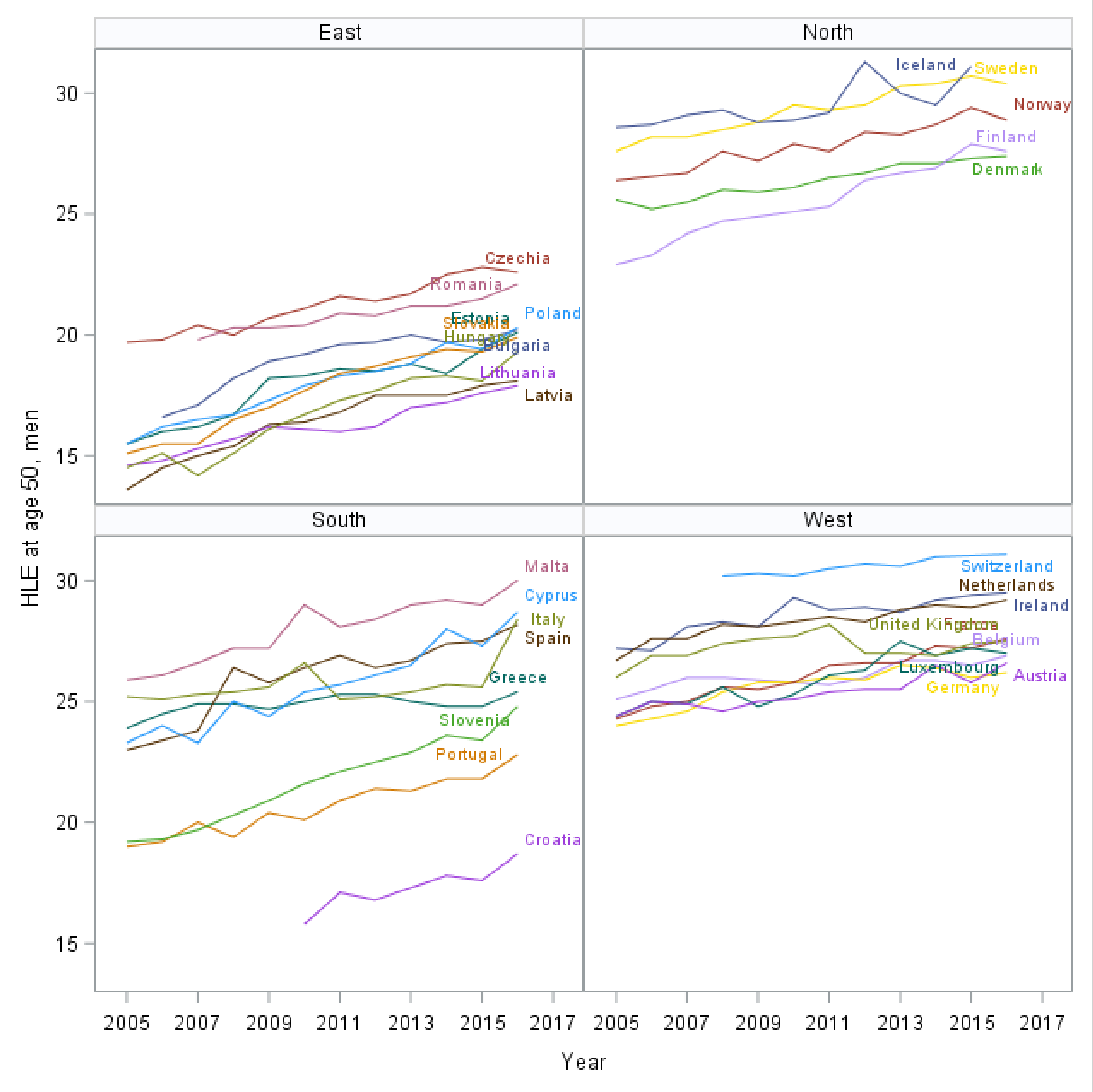
Appendix 6 – Healthy life expectancy based on self-perceived health at age 0 in European countries, men, 2005–2016



Source: Eurostat

Note: processed by SAS 9.4

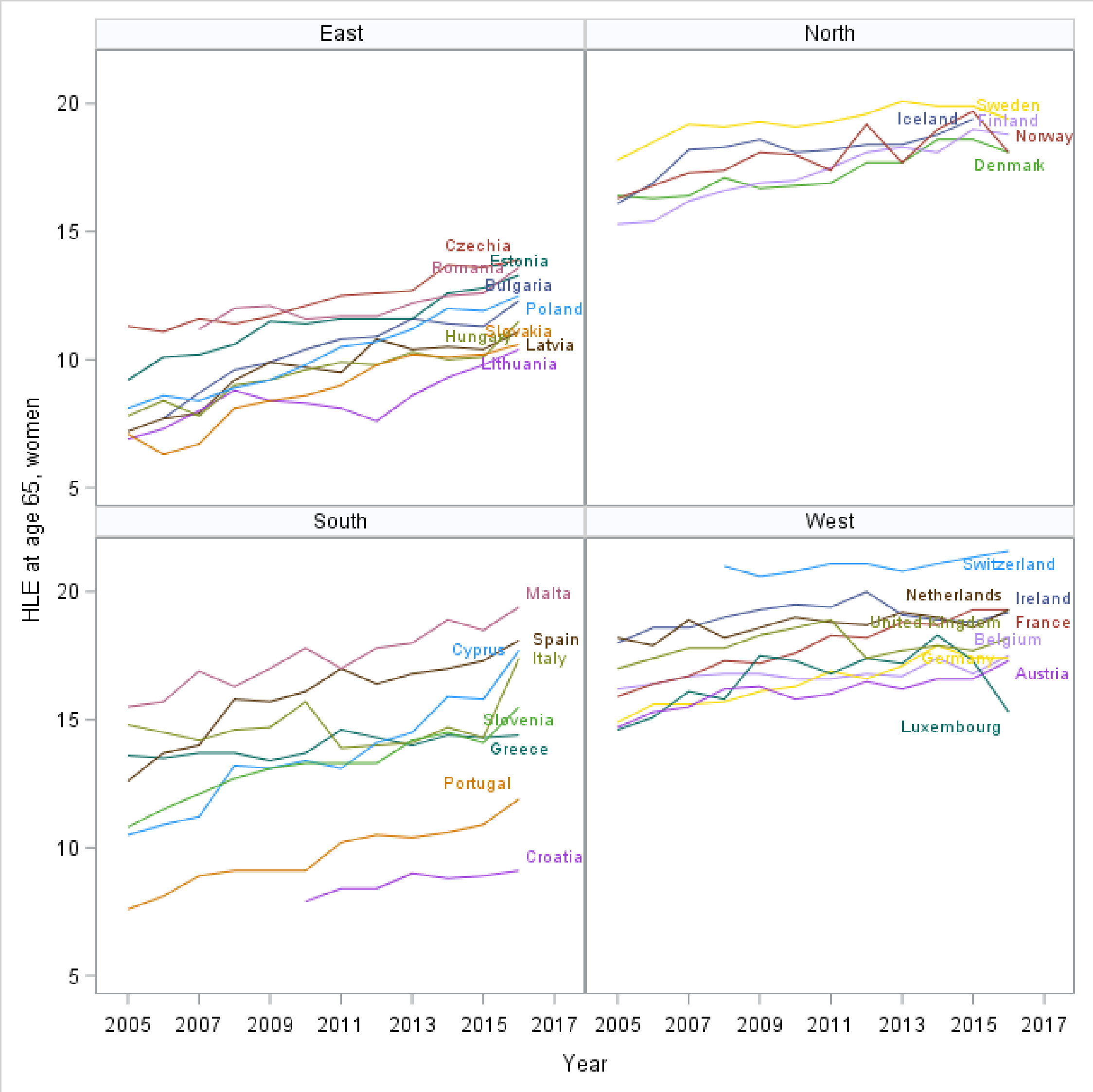
Appendix 7 – Healthy life expectancy based on self-perceived health at age 50 in European countries, men, 2005–2016



Source: Eurostat

Note: processed by SAS 9.4

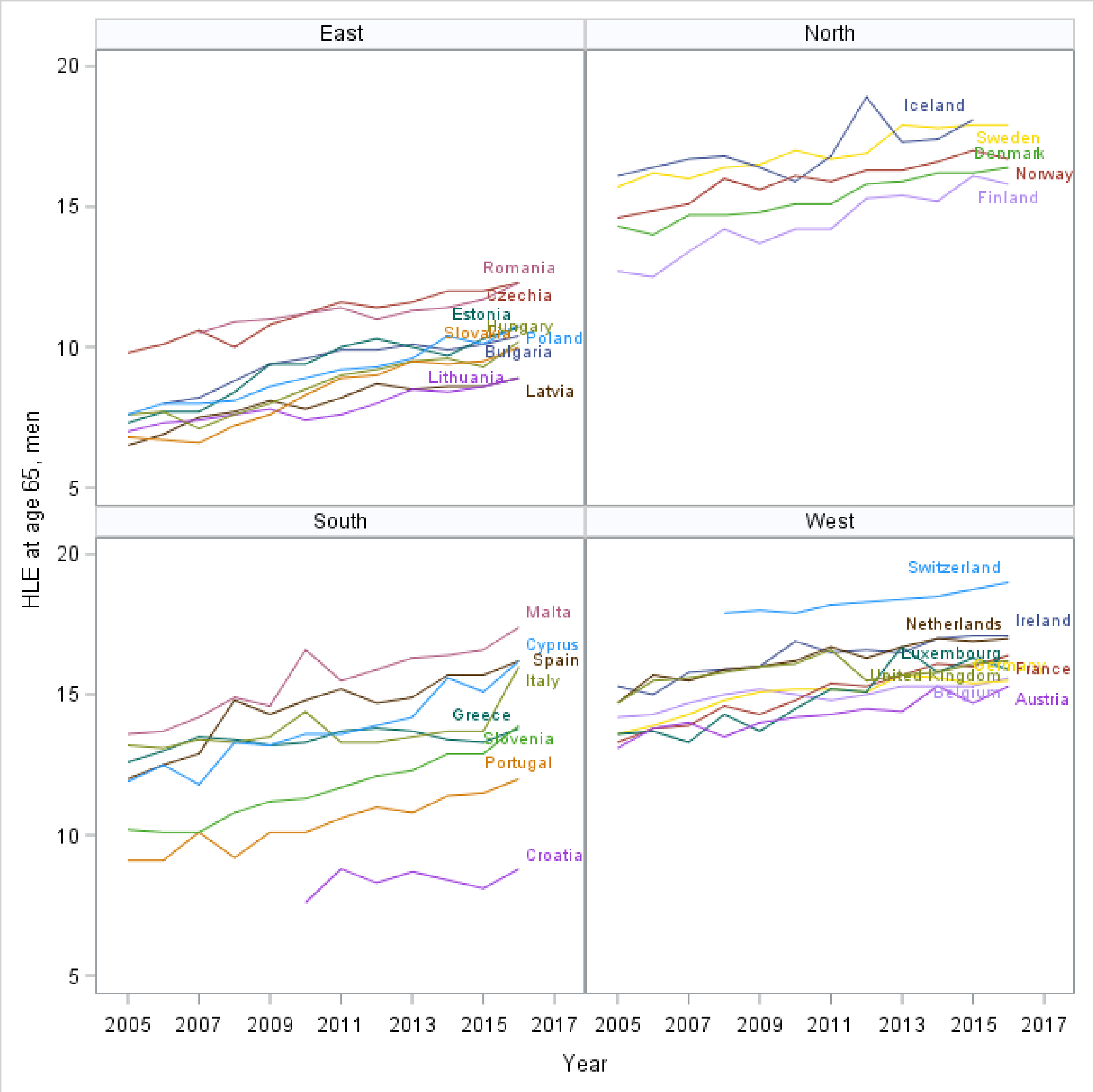
Appendix 8 – Healthy life expectancy based on self-perceived health at age 65 in European countries, women, 2005–2016



Source: Eurostat

Note: processed by SAS 9.4

Appendix 9 – Healthy life expectancy based on self-perceived health at age 65 in European countries, men, 2005–2016



Source: Eurostat

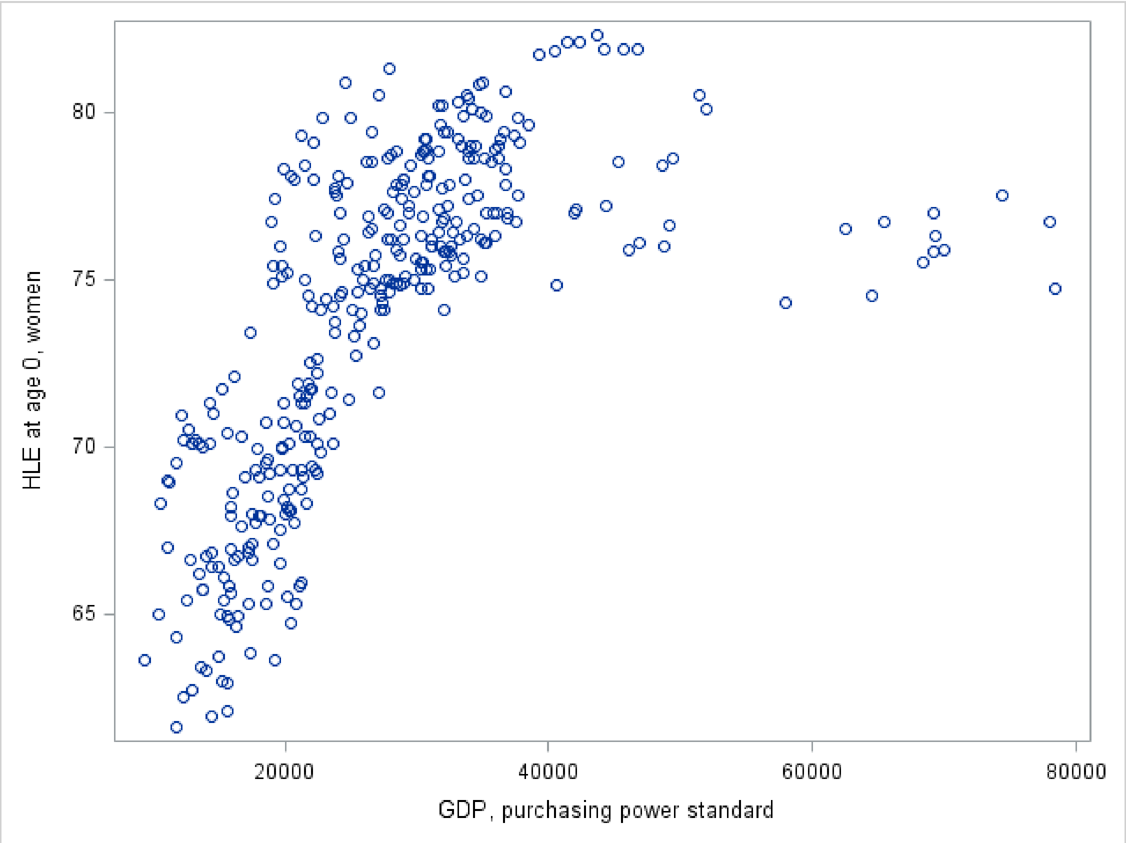
Note: processed by SAS 9.4

Appendix 10 – Pairwise Pearson coefficients of correlation between variables, men (where applicable)

	GDP_PPS	GDP_euro	Hlthexp	Inc_Gini	povrisk_m	povrisk65_m	educ_high_m	educ_low_m	unem_m	unem_long_m
GDP_PPS	1									
GDP_euro	0.97	1								
Hlthexp	0.28	0.36	1							
Inc_Gini	-0.39	-0.41	-0.49	1						
povrisk_m	-0.57	-0.57	-0.58	0.73	1					
povrisk65_m	-0.56	-0.55	-0.58	0.63	0.78	1				
educ_high_m	0.66	0.69	0.18	-0.19	-0.39	-0.36	1			
educ_low_m	0.00	0.00	0.02	0.27	0.08	0.18	-0.20	1		
unem_m	-0.37	-0.37	-0.15	0.43	0.47	0.17	-0.07	0.11	1	
unem_long_m	-0.38	-0.41	-0.17	0.37	0.49	0.16	-0.18	0.08	0.91	1

Source: own calculation based on data from Eurostat

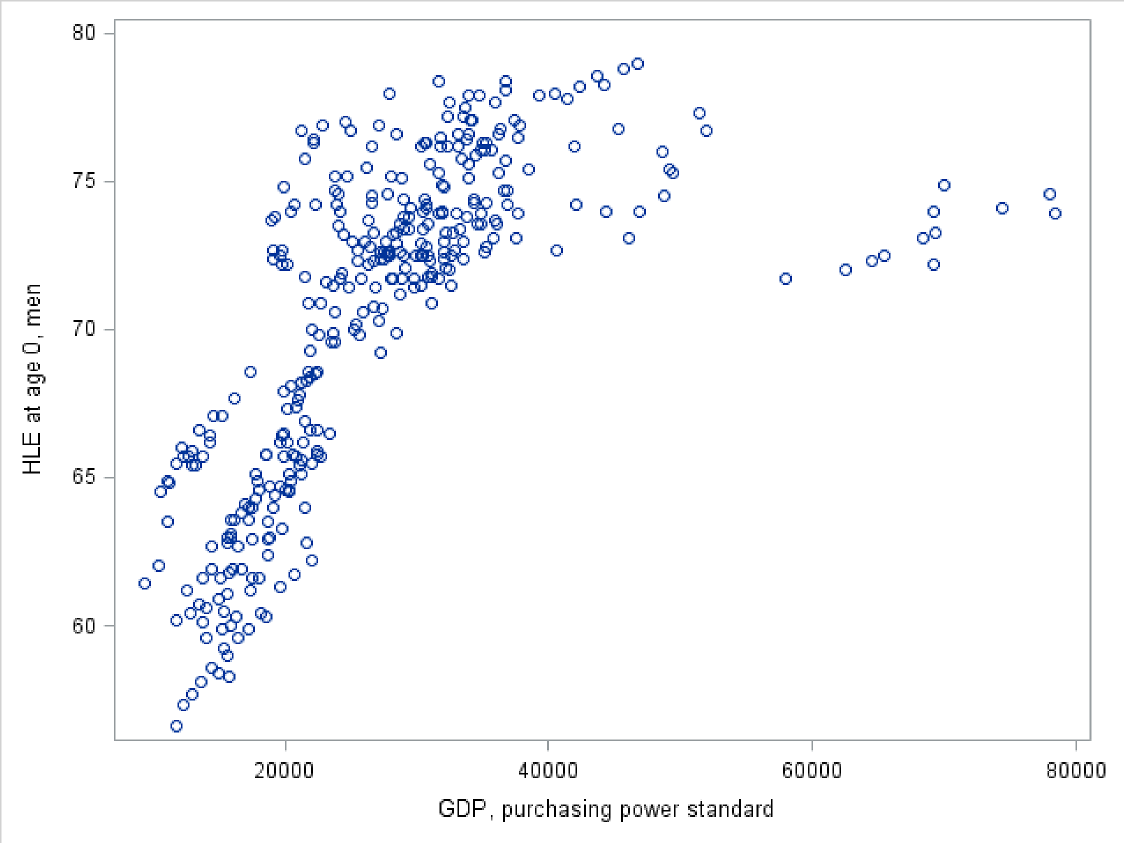
Appendix 11 – Scatterplot of GDP and HLE at age 0 for women



Source: Eurostat

Note: processed by SAS 9.4

Appendix 12 – Scatterplot of GDP and HLE at age 0 for men



Source: Eurostat

Note: processed by SAS 9.4